The Impact Of Faculty Appraisal At Tertiary Level: Two Exploratory Case Studies

Thesis

UNRESTRICTED

JUSTINE ELIZABETH MERCER

(M722422X)

THE IMPACT OF FACULTY APPRAISAL AT TERTIARY LEVEL: TWO EXPLORATORY CASE STUDIES

DOCTOR OF EDUCATION (EdD)

OPEN

JUNE 2004

AUTHOR NO M722422X

The Open University

Faculty of Education and Language Studies

Doctorate in Education

Library Authorisation Form

Please return this form with your bound dissertations.

Student: Justine Mercer **PI:** M722422X

Degree: Doctorate in Education

Dissertation Title: The Impact of Faculty Appraisal at Tertiary Level: Two exploratory case studies

Open University Library Authorisation

I confirm that I am willing for my dissertation to be made available to readers by the Open University Library and that it may be photocopied, subject to the discretion of the Librarian.

Signed: Justine Mercer Dated: 6 August 2004

British Library Authorisation

If you want a copy of your EdD dissertation to be available on loan to the British Library Thesis Service as and when it is requested, you must sign a British Library Doctoral Thesis Agreement Form. Please return it as mentioned above with this form. The British Library will publicise the details of your dissertation and may request a copy on loan from the University Library. Information on the presentation of the dissertation is given in the Agreement Form.

The University has agreed that your participation in the British Library Thesis Service should be voluntary. Please tick either (a) or (b) to indicate your intentions.

(a) ☑ I am willing for the Open University to loan the British Library a copy of my dissertation. A signed Agreement Form is attached.

(b) ☐ I do not wish the Open University to loan the British Library a copy of my dissertation.

Signed: Justine Mercer Dated: 6 August 2004

ABSTRACT

This thesis investigates the impact of faculty appraisal at Rihab University and Al Fanar College, two federal Higher Education institutions in the same Middle Eastern country. It uses a case study approach intended to generate grounded theory from 38 semi-structured interviews, with appraisers and appraisees, alongside documentary analysis, participant observation, and researcher reflection.

The aim of the research is to determine firstly, how far each particular appraisal system is reported to embody paradigms (meaning goals and values, key assumptions, and management ethos) of professionalism and / or managerialism, and secondly, how far each particular system is said by informants to be appropriate to an educational context.

On a more general level, it also looks at the extent to which changes in HE management in 'the West', principally, the United Kingdom, North America, Australia and New Zealand, find resonance in a Middle Eastern context, where HE institutions are staffed predominantly by people from those same Western countries, but are subject to quite different employment laws and practices.

It concludes that although appraisers and appraisees both report a need for appraisal in educational contexts, there is little, if any, correlation between faculty appraisal and improved teaching or learning, most probably because feedback from the process is almost exclusively numerical, and very much oriented towards maintaining minimum standards of technical competency, rather than facilitating individualized, flexible and creative professional development. It suggests that such a state of affairs is not inherent in any appraisal system, per se, but is rather a function of the more general micro-political climate of any particular organisation. That is to say, any appraisal system simultaneously reflects and reinforces the underlying management structures and ethos of an institution.

In the case of both Rihab and Al Fanar, these structures were said to be considerably more authoritarian than would normally be expected in Western countries, for a variety of reasons, including an acute lack of trade unions, tenure, industrial tribunals and legal safeguards. It therefore seems reasonable to conclude that the considerable antipathy many appraisees displayed towards the appraisal systems at Rihab and Al Fanar was more the result of these contextual factors than the particular procedures adopted by each institution. In other words, evaluative appraisal was not seen as automatically undermining pedagogy or professionalism, a very common complaint in much of the previous literature. Instead, it was seen as offering a potentially useful strategy for enhancing educational practices, but only in contexts where employees are protected from the misuse of management power by a variety of legal constraints.

DOCTORATE IN EDUCATION

THE IMPACT OF FACULTY APPRAISAL AT TERTIARY LEVEL: TWO EXPLORATORY CASE STUDIES

THE IMPACT OF FACULTY APPRAISAL AT TERTIARY LEVEL: TWO EXPLORATORY CASE STUDIES

Chapter 1: Introduction

1.1 Focus of the research

The research documented below investigates the impact of faculty appraisal at two Higher Education (HE) institutions, both located in the same Middle Eastern country. It has been motivated by two beliefs. The first is that the faculty appraisal system of an educational institution is particularly indicative of its more general approach to management; in other words, "the kind of appraisal system which an institution adopts reflects and reveals both the value system and the internal structures of the organisation" (Hutchinson, 1997:166). The second is that HE management, in many countries, has experienced an unprecedented degree of change, in the last ten to fifteen years. In other words, appraisal in HE is worthy of study because it is *potentially* "one manifestation of a changing approach to the management of HE (and other state-funded systems) ... across a whole range of countries" (Hellawell, D., personal communication, 3 November 2001).

It is important to note that my research is not about 'appraisal' per se; it does not seek to identify the features of a 'good' appraisal system; it is not concerned with the factors that contribute to the successful implementation of a new system; it is not about the management of change. A great deal has already been written on such issues, and my work adds little to the debate. Instead, it focuses on how HE staff perceive appraisal, and analyses the connections they make between the emergence of more formalised appraisal systems and the call for greater accountability in HE management.

Specifically, my research investigates the systems of faculty appraisal at 'Rihab', a newly-established university, and at 'Al Fanar', an older, more established vocational college, in order to determine firstly, how far each particular system embodies paradigms (meaning goals and values, key assumptions, and management ethos) of professionalism and / or managerialism (as defined below), and secondly, how far each particular system is seen by informants as appropriate to an educational context.

On a more general level, it also looks at the extent to which changes in HE in 'the West', principally, the United Kingdom, North America, Australia and New Zealand, find resonance in a Middle Eastern context, where HE institutions are staffed predominantly by people from those same Western countries, but are subject to quite different employment laws and practices. According to Grace (1985:3), "assessments and evaluations [of teachers] in education have ... implications for the distribution of power and the principles of social control" and as such need to be located "in relation to wider structural, economic and political frameworks". This is undoubtedly true, and my research therefore includes some discussion of these issues, particularly in relation to the generalisability of my findings. Nonetheless, since it would be unwise to make sweeping 'national' or even 'international' generalisations on the basis of two modest case studies, the main focus of my research remains at the level of the institution.

The research began with a pilot study in 1999, and was conducted over a period of four years, using two different sites. During that time, the focus of my research was reshaped and refined many times. What began as a rather vague investigation of perceptions about faculty appraisal ended, four years and 38 interviews later, as a systematic exploration of seven quite specific research questions. For the sake of clarity, I have decided to include these seven questions in the introduction of my thesis, but let me reiterate that the table below represents the final product of a complex process that unfolded over many months in ways I did not always foresee.

1. *How is appraisal, in general, perceived by faculty and managers? How far are accountability models acceptable to either group?*

2. *Is appraisal perceived to have any effect on the quality of student learning?*

3. *What connections, if any, do teachers make between student evaluation of teachers and grade inflation?*

4. *How much time do faculty and management devote to appraisal? Do they consider that the time devoted to appraisal is time well-spent?*

5. *To what extent is appraisal said to focus on either basic technical competencies or flexible, creative practices?*

6. *How far is the specific appraisal system at each institution perceived as developmental and / or evaluative? How far is it seen as embodying a paradigm of professionalism and / or managerialism?*

7. *To what extent do teachers suggest they feel under surveillance?*

Table 1.1

1.2 Claim to originality

I feel that the work I have undertaken is original in two ways. Firstly, as far as I am aware, no previous study of HE in the Middle East has involved more than one research site. By comparing my findings at two similar

institutions, I have been able to make more firmly grounded "fuzzy generalizations" (Bassey, 1999:62) than might otherwise have been the case. Moreover, by analysing my experience of insider research at both places, I have gained a unique insight into the advantages and disadvantages of this research strategy, not least because I have been able to learn from mistakes made at my first site; experiment with alternatives at my second site; and then critically re-appraise both processes. Accordingly, my methodology chapter is rather longer and more detailed than is normally the case, because it is intended not merely to outline how my research was conducted, but also to make a substantive contribution to the field of research methodology, having a significance at least equal to that of my findings and tentative conclusions.

1.3 A conflict of interest?

All researchers are intimately connected with their research, but in this instance, one aspect of my own personal history has had a somewhat greater impact on my work than is normally the case. For this reason, I have chosen to include here some biographical details I believe are particularly relevant to my study. I began investigating faculty appraisal at Rihab in January 1999, more-or-less out of idle curiosity, and the need to undertake a pilot study in order to pass a postgraduate course in research methodology. Having become intrigued by the apparent contradictions inherent in people's perceptions of appraisal, I continued the research for another eighteen months, recording field notes, analysing official documents, and conducting nineteen interviews with faculty colleagues and managers, between February and October, 2000. Entirely unexpectedly, I was informed on 13 December 2000 that my own three-year teaching contract would not be renewed, for reasons that were never made known to me. This decision would have been made in early December, and neither I nor any of my interviewees had any prior knowledge of it. As soon as it happened, I abandoned any further interviewing, and although I continued to record field notes and write entries in my research diary, I subsequently deleted these data, believing them to be excessively biased. Between

January and April, 2001, I began to analyse the data I had collected from my first site, trying, as far as I was able, to avoid unjustified negativity. I restarted my data collection in September 2001, at a different site, where no-one had any knowledge of my previous appraisal experience. These data were analysed between June and October 2002. Finally, I re-analysed the data from my first site, beginning in November 2002. I felt this was a necessary corrective to the earlier analysis I had undertaken, and hoped that the passage of time would have dulled the strength of my emotions and increased the impartiality of my judgement.

I have endeavoured to lay bare my own biography; it is now up to the reader to decide whether or not I have been successful in my attempt to overcome the inevitable degree of bias in my position, and whether or not my research can be of value to a wider audience.

1.4 The organisation of the thesis

Following on from this introduction, I have included a background survey of the literature on appraisal in education, beginning with an examination of such issues as the purpose of appraisal; the alleged tension between professional development and evaluation; and the apparent dichotomy between individual desires and institutional goals. I then go on to explore appraisal in different parts of the world, particularly the UK, the USA, and the Middle East. The last section of this chapter deals with the wider context of appraisal, and considers the various ways in which previous research has linked appraisal to a global rise in managerialism.

It should be stressed at the outset that this survey of the literature is very much an exploration of the background themes that informed the data-collection and analysis stage of my research. It is meant to contextualise my work within a broader framework, and highlights those areas I thought had potential relevance at the time I embarked upon my data collection. It should be seen very much as a precursor to the more detailed and more specific comparisons with the literature found at seven points in chapter

six, where I report my findings in the form of tentative answers to my seven research questions (see table 1.1 above). I am aware that it is more usual for doctoral work to contain a single literature review chapter that both locates the research within the more general field, and addresses particular themes arising from the data analysis stage. I have chosen to follow a slightly different convention because I feel that separating general issues from specific points of comparison is a more effective way of locating my own study within the wider body of research, since it avoids the need to repeat the same material in different chapters.

Chapter three of the thesis outlines my methodology, and contains a detailed discussion of the benefits and drawbacks of insider research, based on my own personal experience. Chapter four gives further details of the way in which the research questions and the interview schedules influenced each other, and evolved over time in a two-stage process. Chapter five discusses the tension between providing "thick description" and protecting participants' confidentiality and anonymity. In light of this, some background information is offered about the two sites and the various appraisal instruments used at each. Chapter six presents the research findings in the form of tentative answers to the seven specific questions. The data from Rihab and Al Fanar are presented separately, and then compared point-by-point with the literature and previous empirical studies. Further details are also given about the wider structural, social and political framework of HE in the Middle East, in an attempt to determine how far the findings of the present study might be generalisable to other contexts. Finally, chapter seven sets out the conclusions I feel can justifiably be drawn from the data I have cited. It also highlights some apparent contradictions in perceptions of appraisal that I feel are worthy of further investigation. It ends with a final statement of the contribution of my research to the theory and practice of education, with respect to both appraisal, in particular, and research methodology, more generally.

Chapter 2: Background Literature Review

As I mentioned in the introduction, the chapter below represents a background survey of those themes that other writers have highlighted as significant in relation to the issue of faculty appraisal. It is meant to contextualise my own study, and should be read very much in conjunction with the material cited throughout chapter six, in which I report my findings and endeavour to show exactly how each of the answers to my seven research questions (listed in table 1.1 above) endorses, refutes and / or extends previous research.

2.1 The concept of appraisal

Unfortunately, but perhaps not surprisingly, the terminology surrounding the concept of 'appraisal' is far from clear. As Fidler (1989:191) notes, "there is a huge variety of terms used – performance appraisal, performance review, performance evaluation, staff review, staff reporting and more especially teacher appraisal, teacher assessment – which have no accepted difference of meaning".

For Magennis (1993:238), the confusion concerns not just the definition of appraisal, but the whole theoretical framework underpinning it:

> Regarding the definition and purpose of appraisal, the procedures that may be used, what constitutes evidence, the reliability and validity of the instruments, there is little consensus.

Fidler and Cooper (1992:2) define staff appraisal as "the process by which an employee and his or her superordinate meet to discuss the performance of the employee", and whilst this is useful in describing what often, though by no means always, happens, it tells us nothing about the purpose of appraisal, and concentrates on the individual employee rather than the organisation as a whole, which not all writers on appraisal would agree

with. Poster and Poster's (1997:148) definition of appraisal is more helpful in this respect. They write that:

> Appraisal is a means of promoting, through the use of certain techniques and procedures, the organisation's ability to accomplish its mission of maintaining or improving what it provides while at the same time seeking to maintain or enhance staff satisfaction and development.

This emphasis on both the organisation and the individual employee is a common theme in much of the literature, though the extent to which these two elements can, or indeed, should, be equally balanced remains highly debatable, as does the extent to which the same scheme can or indeed should be both evaluative and developmental.

2.2 The purposes of appraisal: twin tensions

2.2.1 Professional development and evaluation

Stewart and Stewart (1977), cited in Fidler and Cooper (1992:8), list nine goals of appraisal, ranging from "manpower audit" to "grievance and problem detection and handling". Most other writers confine themselves to somewhat fewer goals, and in almost all cases, these can be subsumed under the two main headings of professional development and evaluation / accountability.

Thus, for example, Beer (1986), cited in Fidler and Cooper (1992:14), suggests three evaluative purposes of appraisal (giving feedback; developing data for pay and promotion decisions; and providing a way to warn unsatisfactory performers), alongside five developmental ones (counselling / coaching; developing commitment to the organisation; motivating subordinates; strengthening supervisor-subordinate relations; and diagnosing individual or organisational problems).

Similarly, Stake (1989:13) writes of four possible purposes of appraisal, the first two being broadly evaluative (providing data about good and bad teaching, and helping managers select and retain the best staff), and the second two being broadly developmental (facilitating continuing professional development and contributing to an understanding of how the school operates).

Likewise, Fletcher (1996:235) gives three possible purposes of appraisal, the first and third being largely developmental, and the second being largely evaluative:

> The typical appraisal system tries to act as a vehicle for motivating people and improving performance (through objective setting), as a means of assessing performance and distributing rewards equitably (through the use of performance related pay and ratings) and as a development tool ... .

Turner and Clift (1988:9) focus appraisal somewhat narrowly on the teacher, but here again, there is still evidence of this twin emphasis on development (helping teachers improve) and evaluation (deciding on probation, renewal and dismissal). Interestingly, Turner and Clift (1988:52) go on to suggest that developmental appraisal is "formative", whilst evaluative appraisal is "summative" - a parallelism that is far from self-evident, since evaluation could be on-going and development short-lived.

How far it is desirable, or even possible, to combine these two elements of appraisal within a single institutional system, and particularly within the same time-frame, remains a moot point. For Poster and Poster (1997:152), appraisal has no value unless it brings together both staff development and performance review. Likewise, Turner and Clift (1988:20) contend that formative and summative appraisal (meaning developmental and evaluative) "are not easily divorced and it is likely that any scheme will be adopted to some extent for both purposes". Nuttall (1986:23), cited in

Campbell (2002:11), agrees, claiming that "formative and summative purposes can co-exist within the same scheme".

There is also a certain amount of empirical evidence to suggest that teachers and administrators see evaluation and development as complementary rather than contradictory forces. In Peaker's (1986) case study of two Southern USA school systems, teachers, principals and administrators all saw appraisal as having two functions ("improvement" and "accountability") that were "entirely compatible, even interdependent" (Peaker, 1986:78). Similarly, Fitzgerald (2001:120), cited in Campbell (2002:12), reports that 99% of the New Zealand schoolteachers in his survey agreed that appraisal should include elements of both professional development and accountability. For Middlewood (2001:131-132), cited in Campbell (2002:12), the figure amongst schoolteachers in the British Midlands is 90%.

Interestingly, the writers of many non-empirical articles disagree. Winter (1989b:48), for example, claims that "there is a readily available body of evidence and argument suggesting that a plurality of aims renders appraisal systems ineffective". Prominent within this body are Maier (1958) and Duke and Stiggins (1985), both cited in Hellawell (1991), Fidler and Cooper (1992:3), Powney (1991b:84), Fletcher (1996:235), Sale (1997:29), and Taylor (1997:56).

Fidler and Cooper (1992:3) are unequivocal when they write of "the fundamental contradiction inherent in using appraisal for both evaluative and developmental functions". Similarly, Lovrich (1990:99), cited in Ko (2001:27-28), describes evaluation and development as "two very different – and in many ways antithetical – phenomena". Likewise, according to Sale (1997:29):

> It cannot be stressed enough that to confuse the two activities of development and assessment under one umbrella of appraisal is fatal to staff development and arguably inefficient and unfair as a means of assessment.

The same sentiment is expressed by Powney (1991b:84):

> Appraisal must be either about development or about judgement. It must be either formative or summative. It cannot be both.

This is because:

> The relationship between appraiser and appraisee for goals such as developing data for pay and promotion decisions appears to contradict the trusting relationship required for counselling appraisees in areas of organisational commitment and future development.
>
> (Taylor, 1997:56)

In other words, "The imminence of reward decisions tends to block constructive discussion of developmental needs" (Fletcher, 1996:235).

One way round this problem is to separate the two aspects of development and evaluation by at least six months (Fidler and Cooper, 1992:16), and use different instruments at different times with different people (House and Lapan, 1989). In this way, summative assessment using a "craft model" of teaching, with its emphasis on "a repertoire of specialised techniques and knowledge", would be used with all teachers occasionally, to maintain minimum standards of competency; but the main focus would be on formative assessment, using "professional" and "art" models of teaching, with their emphasis on "the exercise of judgement" and "personalised" teaching in the face of the "unpredictable" (House and Lapan, 1989:56-60).

2.2.2 The individual and the institution

As we saw earlier, the tension in appraisal between evaluation and development in mirrored by a similar tension between the needs of the individual and the demands of the institution. And just as opinion differs as to how far evaluation and development are compatible, so too is opinion divided over how far the needs of the individual and the demands of the institution can be matched. Poster and Poster (1997:152) saw little, if any,

tension between evaluation and development, and, not surprisingly, their position is the same with regard to any potential tension between the individual and the institution: "Appraisal is one of a number of procedures for integrating the individual into the organisation ... Appraisal must be for the benefit of *both* the individual *and* the organisation" (Poster and Poster, 1997:152, italics in the original).

Unfortunately, the reality does not always match Poster and Poster's ideal. Very often the focus of appraisal is either individual professional development, without reference to the goals envisioned by the institution; or else individual performance evaluation, without reference to the constraints imposed by the institution.

With regard to the former scenario, Bennett (1999:415), cited in Campbell (2002:81), lists several major weaknesses of appraisal in schools, including a failure to ensure that appraisal is linked to school development plans. Just such a failure is evident in the study of headteacher appraisal conducted by Cullen (1997:181), where the heads "clearly viewed appraisal as serving to promote their personal, professional development" without any mention of "the larger aim of improving the quality of educational provision in schools".

With regard to the latter scenario, Fletcher (1996:239) cites Deming's (1986) suggestion that "performance appraisal is one of the seven deadly diseases of current management practice" because managers do not differentiate between individuals and the organisational systems within which they work.

Simons and Elliott (1989:13) are aware of this pitfall, claiming that:

> The evaluation of teaching is an inseparable part of the evaluation of the school, itself. An effective evaluation of teaching requires concurrent study of institutional goals, classroom environments, administrative organisation and operations, curricular content, student achievements, and the

impact of school programmes on the community.

Similarly, Walsh (1988:365) contends that it is pointless to appraise the individual without reference to the conditions under which she works. If appraisal were used "not to identify the successes and failures of the individual teacher, but the constraints that operated on the school as a learning community", and if it involved a wider variety of people, such as parents and pupils as well as teachers and heads, it would promote "a reciprocal form of accountability, rather than a hierarchical, managerial accountability" (Walsh, 1988:368).

Yet, as both sets of authors concede, such an integration of the individual and the institution rarely happens because all too often the appraiser is also the manager partly responsible for the creation of the conditions under which the appraisee must work. In these circumstances, it is easier to blame the incompetence of the teacher than share responsibility for the inadequacy of the environment, especially when this includes a lack of good management.

2.3 Appraisal around the world

2.3.1 UK

Whilst a great deal has been written about appraisal in Britain, not many empirical studies have been conducted, and those that have have tended to concentrate on the government-school sector rather than HE. Thus, for example, Kyriacou (1995) interviewed forty teachers at different schools within one Local Education Authority and concluded that "most schemes of teacher appraisal adopted have emphasised a professional development model and have generally been well-received" (Kyriacou, 1995:109). The vast majority of teachers said the experience was positive, particularly the interview stage, and added that it was valuable to have time set aside specifically for reflection and discussion. They appeared to appreciate the opportunity to get feedback from a colleague and set negotiated targets.

About two-thirds of the teachers reported that the process had led to changes in their classroom practice. With regard to disadvantages, the most common complaint was about how much time the process took, and whether pupils might suffer from the amount of extra cover required. There was also some concern about whether the resourcing of action plans was adequate.

These findings concur with those of Gunter (1996), who used a questionnaire with thirteen staff at the same school, and reported that they had made "appraisal a positive process which is perceived to have some impact on professional practice" (Gunter, 1996:99). All respondents thought the initial meeting, self-appraisal, and data collection through classroom observation were helpful or very helpful, and over 70% of them (presumably 10 out of the 13) also found the post-data-collection meeting, target setting, and document completion helpful or very helpful. Nonetheless, the impact of appraisal was being hindered by a lack of time and resources, as well as a fragmented approach that emphasised task-completion, rather than personalised processes. Although teachers felt positively about appraisal, it was not helping the institution become a learning organisation because its impact was short-term, and mainly focused on classroom observation, and because, in most cases, targets were not being monitored.

The teachers in Powney's (1991a) study also found appraisal a positive experience, though many emphasised that this was because the system adopted was formative / developmental rather than evaluative: "There is a clear commitment to the formative or developmental approach and a marked antipathy to appraisal as a simplistic judgement of teacher performance" (Powney, 1991a:185). Interestingly, the teachers in this study occupied 'middle management' positions and were keen for the appraisal process to examine all facets of their role, particularly how well they managed other people. Whilst Gunter (1996) thought the appraisal system she studied had not changed the school into a learning community, the middle managers in Powney's (1991a) study were more hopeful:

> Most of the teachers interviewed were of the opinion that a properly run, formative appraisal system, possibly using the department or team as the appraisal group, could help to foster the climate of security and openness in which aims are cooperatively formulated and change can occur and be examined on its merits.
>
> (Powney, 1991a:187)

By contrast, the headteachers in Cullen's (1997) study saw their own appraisal as having little effect on the quality of teaching and learning, even thought they found it valuable in other ways. They reported that it made them feel better about their job because they received positive reassurance and feedback, and it enabled them to perform better because they were able to take stock, obtain an outside view from the head at a similar school who appraised them, engage in professional dialogue, and develop relationships. Having staff see their head being appraised also gave the system credibility. With regard to the disadvantages, headteacher appraisal was seen as logistically difficult to organise, very time-consuming, and inadequately funded. There was also concern that appraisal was a low priority in the life of the school, and could easily become superficial if the focus were not well-chosen, or the appraiser too afraid to offer constructive criticism.

Turning now to appraisal in Higher Education, Rutherford (1988) found from his survey of academic staff that most faculty accepted the need to appraise individual performance, as long as this was heavily contextualised. In other words, there was:

> ... widespread acceptance that further systematic procedures for the appraisal of individuals are necessary; that appraisal should be as comprehensive as possible; and should include an annual interview with the Head of Department .... It seems sensible that arrangements for appraisal should be kept as flexible as possible to take account of local contexts

> and personalities; centrally imposed solutions are unlikely to command wide acceptance.
>
> (Rutherford, 1988:98)

Taylor's (1997) survey of HE faculty was reported in more detail than Rutherford's (1988), and because it distinguishes between appraisers and appraisees, and was conducted in two phases over two years, it provides some very interesting data about differences in perception over time. The areas of investigation included the topics discussed in the appraisal interview, goal-setting, peer appraisal versus supervisor appraisal, performance related pay (PRP), observation of teaching, and student evaluations.

In both years, three of the six appraisers said the appraisal interview did not cover important areas worthy of discussion, though few of the appraisees agreed with this, particularly in the second year. Thus it would seem that the appraisees were happier with the content of the discussion than the appraisers, which seems strange, since one would have expected the appraiser to have had greater control over the choice of interview topics. With regard to goal-setting, some of the appraisees reported feeling pressured into setting unrealistic goals, although none of the appraisers seemed aware that they were doing this. A third of the appraisees said they would prefer peer rather than supervisor appraisal, although, again, none of the appraisers agreed. The number of people in favour of PRP increased in the second year, but was still confined to 33% of appraisees and 50% of appraisers. Finally, with regard to observation and student evaluation, 45% of appraisees said observation was valuable, whereas 58% said student evaluation was valuable. With appraisers, the position was reversed, with only 50% saying student evaluation was important, but 83% saying observation was important. These results point to some significant differences in the reported perceptions of appraisees and appraisers, suggesting that appraisees may have a more acute sense of the power dynamics inherent in the process.

These differences in perception persisted when respondents answered two open-ended questions. When asked to comment on the benefits of a good appraisal scheme, appraisers focused on appraisal as a mechanism for motivating staff to work harder, whereas appraisees focused on identifying personal goals and planning long-term development. When asked about the barriers to achieving a good appraisal scheme, both appraisers and appraisees pointed to a lack of commitment to the process at senior management level, but the appraisees also mentioned inadequate resourcing and poor interpersonal skills on the part of the appraisers.

To summarise the findings from all sectors, it seems that appraisal in the UK is seen as a potentially positive experience and one that will enhance professional development, if not teaching and learning directly, as long as the focus remains formative, and full account is taken of contextual factors. There is some concern about the amount of time appraisal takes, and about whether or not the action plans deriving from it will be adequately resourced. To be successful, appraisers are thought to need highly-developed interpersonal skills and specific training for their role.

2.3.2 USA

When comparing Britain and the USA, most commentators argue that appraisal schemes in the USA have a more evaluative, competency-based focus, with a greater emphasis on classroom observation checklists and measurements of pupil achievement. See, for example, Turner and Clift (1988:10), Powney (1991b:84), Mortimore and Mortimore (1991:125), Fidler and Cooper (1992:43), Richards (1999, 5-6), and Campbell (2002:16). The "deficit model" (Campbell, 2002:19) of professional development seems more in evidence, bringing with it a concern to make sure that all teachers are at least adequate, rather than a desire to help them move from adequate to good, or even good to excellent.

House and Lapan (1989:56-57) suggest that most appraisal schemes in the USA have a "craft" perspective on teaching, which, as we saw before,

includes an emphasis on acquiring and demonstrating "a repertoire of specialised techniques and knowledge":

> Teachers are expected to carry out their duties without close supervision and the role of the administrator-evaluator is to hold them to performance standards by visits, conferences and written agreement.
>
> (House and Lapan, 1989:57)

This focus on competency-based evaluation can also be seen in Peaker's (1986) empirical study of teacher appraisal in two school systems in the Southern USA. Peaker (1986:83) claims the appraisal procedure he studied required teachers to demonstrate highly specific competencies and, as such, may have encouraged safe, didactic teaching, at the expense of more flexible, creative teaching. He also reports an emphasis on what is easily measurable, such as test scores, rather than a more holistic assessement of the teacher's overall contribution. On the other hand, most of the teachers and administrators themselves viewed the appraisal system as a success, and Peaker attributes this to the high levels of training and staff development resourcing.

2.3.3 The Middle East

I was very fortunate in being able to track down a case study on appraisal by Richards (1999), conducted four years ago at a sister college in the next county to Al Fanar. This college is part of the same federally-funded system of HE, offering very similar educational provision, being governed by the same statutes and bye-laws, and having more-or-less the same management structure as Al Fanar.

The case study was based on the results of a questionnaire sent to all seventeen members of the English department, and six interviews, two with managers and four with faculty. The research focused on a number of issues including the perceived purpose of appraisal; the appraisal instruments (student evaluation, classroom observation and teaching

portfolios); the appraisal process (goal-setting and discussion), and the appraisers, themselves.

Regarding the purpose of appraisal, the two managers saw it as both formative and summative, without any inherent conflict. The teachers, however, thought the aims were ambiguous or unclear, and in some cases, this created suspicion and confusion. Whilst the managers stressed the need to have consistent, fair and objective information about teachers, the teachers, themselves, remained concerned about exactly how the information might be used, one of them even suggesting that it could be used negatively as the basis for dismissal, but not positively as the grounds for continued employment (Richards, 1999:41-42).

With regard to the appraisal instruments, 23% of teachers were dissatisfied with the system of student evaluation, and 38% were very dissatisfied. Several teachers thought many students could not assess a teacher's ability objectively; others complained about the timing, the content, or the specific wording of the feedback form (Richards, 1999:49-50). By contrast, there was a high level of satisfaction with all aspects of the classroom observation, including the pre-lesson briefing and the post-lesson debriefing. In this system, it seems that teachers were able to choose from a variety of observation models, which included one complete lesson visit, two half-lesson visits, or several 10-minute drop-ins, a choice that allowed managers and teachers to address some of the obvious 'sampling' problems that accompany a single observation. Almost all teachers reported that the observation process had been a positive experience that improved their morale and motivation, though not necessarily their teaching. Richards (1999:72) suggests that this is more a reflection of the particular appraiser's perceived skills in observation and interpersonal communication than inherent in the system itself. Opinion was again more mixed regarding the teaching portfolio, with which 31% of faculty were dissatisfied, and 15% very dissatisfied. Teachers were concerned that it may become overly competitive, or that people who were better writers than teachers would benefit unfairly. Some teachers said they found the requirement to reflect

on practice helpful, whilst others said they found it unnecessary, suggesting that "reflective practice as a basis for staff development is not necessarily applicable or acceptable to all teachers" (Richards, 1999:60).

Concerning goal-setting, the college was reported to have adopted a three-stage process whereby the teacher set her own goals, at the beginning of the year, then agreed them with her line-manager, and reported back, at the end of the year, on the extent to which they had been achieved. Most teachers commented positively on the usefulness of this approach (Richards, 1999:57).

They also commented very positively on the end-of-year appraisal interview, because it was formative, rather than summative, and because the appraiser concerned was perceived to be very supportive and very much in favour of the whole process (Richards, 1999:62-63). Two teachers did raise the question, however, of whether entirely positive feedback was effective, and whether appraisers would be honest in cases of poor performance.

Overall, both the managers and faculty in Richards' (1999) case study seemed to view appraisal in a positive light. A large part of this, Richards (1999:72) suggests, is down to the ability and enthusiasm of the particular appraiser, who was well-trained and experienced in appraisal procedures, and well-respected by the teachers he appraised. Accordingly, Richards (1999:72) makes a strong case for appraiser training, particularly if the appraiser has recently moved into education management from a non-teaching background in, say, industry or business.

2.4 Appraisal within the wider structural, economic, and political framework

As I said in the introduction, the main focus of my research is appraisal at the institutional level, but bearing in mind Grace's (1985:3) injunction, cited earlier, to locate evaluation systems within their "wider structural,

economic and political frameworks", I intend to end my background literature review with some discussion of the global rise in managerialism. A great many commentators endorse Walsh's (1988:367) contention that "The development of appraisal systems is part of the development of a more managerial approach to education", and for this reason, the concept of managerialism is worthy of further exploration, even though, ultimately, it did not figure as prominently in the minds of my interviewees as the literature had initially suggested it might.

2.4.1 A plethora of terms

We saw earlier how the concept of 'appraisal' is understood and described by different authors in quite different ways. The situation is similar, if not even more complicated, with regard to recent trends in the management of public sector institutions. Different authors have used different terms to describe these trends, and it is not always clear whether different authors are using different words to describe the same phenomenon, or, worse, whether different authors are using the same words to describe different phenomenon.

Thus, for example, the term "managerialism" is used by Inglis (1989), Enteman (1993), Trow (1994), Webb and Vulliamy (1996), Woods and Jeffrey (1996) Clarke and Newman (1997), Kydd (1997), Trowler (1998a), Shelley (1999), Currie and Vidovich (2000), Lumby and Tomlinson (2000), and Simkins (2000), whilst the term "new managerialism" is used by Hartley (1997), Deem (1998), and Exworthy and Halford (1999). In Hartley (1997), this "new managerialism" is contrasted with the old managerialism of the late 19th century and the Enlightenment, whereas, in Exworthy and Halford (1999:7), it is contrasted with the Taylorist ("bureaucratic, inflexible, conformist") management of the early 20th century. After a careful reading of all these texts, it seems safe to conclude that "managerialism" and "new managerialism" are used synonymously in the literature. Further support for the interchangability of these two terms can be found in the fact that Randle and Brady published two articles in the same year (1997), using the same data set, the first entitled "Further

Education and the New Managerialism", and the second entitled "Managerialism and Professionalism in the 'Cinderella Service' ".

Whether the term "New Public Management" (NPM) is also synonymous with "managerialism" is less clear. Holloway (1999) seems to equate the two, whilst Peters et al. (2000) describe NPM as "a specific constellation of these theories", previously referred to as "new managerialism". For Pollitt (1993), however, "managerialism" is not the same thing as NPM at all. In chapter seven of his book, "managerialism" is contrasted with both "Public Service Orientation" (PSO) and "New Public Management", though previously, in chapter six, the contrast was with "Public Service Orientation" and "public management". Thus, it seems that "New Public Management" (NPM) and "public management" are probably the same thing, which is, itself, very different from "managerialism", and somewhat different from PSO.

Interestingly, not all authors agree with Pollitt's typology, particularly his carefully drawn distinctions between "managerialism", NPM and PSO. Some authors appear to conflate all three terms. Thus, for example, Flynn (1999:28) argues that "New Public Management" is sometimes called "Public Service Orientation" and represents a "fusion of best practice" from both the private and the public sector. The essential components of NPM / PSO that he then lists (more active and accountable management; explicit standards, targets and measurement of performance; a stress on results, quality and outcomes; and so on) sound very similar to many other descriptions of "managerialism" and "new managerialism".

And, as if this were not confusing enough, the debate is further complicated by the introduction, in Trow (1994), of the terms "soft" and "hard" managerialism, further details of which will be discussed below.

Now, it could be argued that this enduring confusion over key terminology is indicative of 'woolly' thinking. However, I prefer to see it as the almost inevitable consequence of trying to describe a very complex and deeply

contested vision of reality. The management of public sector institutions has never been so complicated, nor so controversial. However, since the focus of my research is faculty appraisal and its relationship to the management of HE institutions, rather than managerialism per se, the precise details of the various fine-grained distinctions need not overly concern us. For the purposes of my research, the similarities between "managerialism", "new managerialism", "New Public Management" and "Public Service Orientation" are more important than the differences. For this reason, I will use the term "managerialism" in a general sense, and concentrate only on those over-arching features that seem common to all four of the above terms.

2.4.2 The concept of managerialism

According to Deem (1998:47), "The term 'New Managerialism' is generally used to refer to the adoption by public sector organisations of organisational forms, technologies, management practices and values most commonly found in the private business sector". It represents "a way of trying to understand and categorise attempts to impose managerial techniques more usually associated with medium and large 'for profit' businesses onto public sector and voluntary organisations" (Deem, 1998:49). Such techniques include "the use of internal cost centres, fostering competitiveness between employees, marketisation of public sector services, and the monitoring of efficiency and effectiveness through measurement of outcomes and individual staff performances" (Deem, 1998:49-50). A similar list of "management techniques" is presented by Randle and Brady (1997b:125), drawing, as they say, on Pollitt (1990). These techniques include:

> strict financial management and devolved budgetary controls; the efficient use of resources and an emphasis on productivity; the extensive use of quantitative performance indicators; the development of consumerism and the discipline of the market; the manifestation of consumer charters as mechanisms for accountability; the creation of a disciplined, flexible workforce, using flexible /

> individualised contracts, appraisal systems and performance related pay; [and] the assertion of the managerial control and the managers' 'right to manage'.

Deem, and Randle and Brady write from a British perspective, Pollitt from an Anglo-American one. Their descriptions are echoed by others, writing, for example, in Australia (Currie and Vidovich, 2000) and New Zealand (Boston, 1996). These latter writers might use the term New Public Management, rather than "managerialism", but the techniques they describe are unmistakably similar. Thus, for example, Peters et al. (2000:116), summarising Boston (1996), list the features of NPM as:

> an extensive use of written contracts and performance agreements; a reliance on short-term employment contracts; an emphasis on economic rewards and sanctions; a reduction in multiple accountability relationships; a minimising of opportunities for ministerial discretion in the detailed operation of government agencies; the institutional separation of the funding agency from the provider; the separation of advisory, delivery and regulatory functions; an introduction of accrual accounting; capital charging; a distinction between the state's ownership and purchasers' interests; a distinction between outcomes and outputs; an accrual-based appropriation system; and an emphasis on contestable provision and contracting out for service.

It would, therefore, seem that the technical features of managerialism are common to a variety of different countries. Yet, as Trowler (1998a:93) points out, new managerialism is much more than merely a set of techniques:

> To see managerialism as simply a box of tricks is to ignore its ideological essence. Managerialism involves a framework of values and beliefs about social arrangements and the distribution and ordering of resources. This provides a guide and a justification for behaviour.

Other authors concur. For Enteman (1993:154), "Managerialism is both a process and a substantive ideology", whilst for Pollitt (1993:6), it is an ideology, in the sense that it includes beliefs about the state of the world and what the world should look like. It incorporates "beliefs" as well as "practices" (Pollitt, 1993:1); "ideologies" as well as "institutions" (Clarke and Newman, 1997:ix); "theories" as well as "models" (Peters et al., 2000:110).

Central to the underlying ideology of managerialism is the belief that public sector institutions require more rigorous management in order to operate better. "Managerialism promises to provide the discipline necessary for efficient organisation, particularly in relation to state welfare professionalism's claims to exercise discretionary judgement" (Clarke and Newman, 1997:30). It "was presented as the means through which more rigorous discipline could be introduced to the public sector to produce more cost-effective services (and thus limit public spending)" (Clarke and Newman, 1997:34). It assumed that "good management", as copied from the private sector, would deliver economy, efficiency, effectiveness and value for taxpayers' money (Randle and Brady, 1997a:230), something, which, by implication, had been previously lacking.

Thus, from a managerialist perspective, public sector institutions were underachieving, and the application of private sector values and technologies would bring about the required improvement. Naturally, such a perspective did not emerge overnight, though Randle and Brady, (1997a:230) do date it, perhaps a little too precisely, to 1979, the year Margaret Thatcher became prime minister. Nor did it emerge in a vacuum. It came about within the context of certain global social, economic, and political trends. It was "both a source and a beneficiary of ... wider economic and political transformations" (Clarke and Newman, 1997:58). It is to these trends / transformations that we now turn, since, in the words of Pollitt (1993:168) "a grasp of both the socio-economic and ideological shifts is essential to understanding the rise of managerialism".

2.4.3 The historical context: global changes in the public sector

For Pollitt (1993), the driving force behind the rise in managerialism was the dominance of New Right politics in both the US and the UK during the 1980s. "The Reagan and Thatcher administrations both came to power expounding the idea that government had grown too big, too expensive, and too inhibiting of individual enterprise" (Pollitt, 1993:48). Their solution was to privatise as many public services as possible (most notably, the British utilities). The remaining services, such as healthcare, education and the civil service, being either too impractical or too close to the public's heart to privatise, had to be reshaped in order to improve their productivity. Quality had to be maintained, or preferably increased, even though the level of government funding was being reduced. The way to achieve this particularly challenging political objective was through the innocuous sounding "better management".

Exworthy and Halford (1999:1) agree that the policies of successive Conservative governments provide an obvious explanation for what they term a "dramatic transformation right across the public sector":

> Through policies which emphasised a belief in individualism rather than collectivism, in markets as the most efficient distributor of resources and in reducing dependence on welfare state provision, New Right policies have promoted managerialism in the public sector and conceptualised this as part of a broader strategy to deal with (what they believed to be) Britain's pervasive and persistent economic problems.
>
> (Exworthy and Halford, 1999:8)

They also suggest that the New Left supported the rise of managerialism, just as much as the New Right, although for rather different reasons, which explains why many of the Conservative initiatives in the public sector were continued, rather than reversed, by the incoming Labour government, in 1997. Whilst the New Right saw managerialism as a way of improving efficiency, the New Left saw it "as a means of bringing services closer to the users and making them more responsive to their needs" (Exworthy and

Halford, 1999:8). The arguments from both sides of the political divide are neatly summarised by Flynn (1999:19) as follows:

> For those on the New Right (especially the 'public choice' school), state intervention was criticised for undermining economic efficiency and investment, as well as debilitating the enterprise culture through the promotion of welfare dependency. Professional groups were seen as self-serving monopolies whose influence on the economy and society was negative. The New Left also vilified professionals for exploiting clients in private markets, and accused professions in the paternalistic welfare state of disempowering citizens, while facilitating bourgeois domination and social control.

"Striking uniformity" is seen by Pollitt (1993:83) in the application of managerialism to three key UK public sector services, namely, the Civil Service, the National Health Service and education. All three have witnessed reduced funding; cuts in staffing; the development of performance indicators stressing economy and efficiency; the introduction of more formalised individual staff appraisal including performance related pay; more devolved budgetary systems; more management training; more emphasis on short-term, outcomes-based planning; and more rhetoric about responding to the needs of the consumer.

By contrast, other writers have highlighted the differences rather than the similarities in the ways in which managerialism has emerged in different public services and in different subsectors of the same service. "Managerialism is not a unified set of discourses and practices" (Clarke and Newman, 1997:84). Rather, it "is enacted in different ways in different settings" (Clarke and Newman, 1997:99). Not only does it take many forms, it also "interacts with the existing organisational orders in a variety of ways to produce different outcomes in different contexts" (Simkins, 2000:330). Factors influencing this variation occupy a range of positions on the macro / micro-level continuum.

At one extreme, there is the government policy framework affecting the whole sector; next, there are subsector differences. For example, Simkins (2000) argues that FE colleges have faced a more acute funding crisis than schools, and, as a consequence, the need to demonstrate efficiency has been felt more keenly by FE colleges than by schools, who, in contrast, have directed their efforts towards demonstrating effectiveness to Ofsted inspectors.

Then, after subsector differences, come factors concerning "the relative positioning" of a particular institution vis-à-vis the market competition, as well as its "cultural starting point", meaning what the institution was like before the emergence of managerialism (Simkins, 20000:330). Finally, there are the preferences and styles of particular organisational leaders and managers.

2.4.4 The appropriateness of managerialism in educational contexts

There seems to be a certain lack of agreement about how far managerialism is appropriate to the public sector, in general, and to educational contexts, in particular; a divergence of opinion that seems to rest upon the extent to which modern managerialism and traditional professionalism are seen as either contradictory or complementary paradigms.

Exworthy and Halford (1999:125) suggest that professional-managerial relations can be understood in three ways, namely abstractly, collectively, and individually. The analytical or abstract level deals with ideological or theoretical underpinnings; the collective level deals with the collective body of managers or professionals; the individual level deals with a single human being and their fluctuating self-identity. Exworthy and Halford (1999) argue that those who emphasise the co-existence of managerial and professional roles within the same person are operating at the collective / individual level, while those who emphasise irreconcilable differences in values are operating at the analytical level. In this way, they are able to account for the apparent diversity of opinion that abounds when different

writers discuss the relationship between managerialism and professionalism.

Alternatively, one could account for the differences by invoking the distinction, first made by Trow (1994), between "hard" and "soft" managerialism. According to Trow (1994:11), "soft" managerialists believe that HE should remain governed by its own norms and traditions, but that the goals the academic community sets itself can be achieved more effectively through the adoption of some (primarily "rational") business practices. By contrast, "hard" managerialists think that HE is not fit to set its own agenda and should be reformed and controlled by funding formulas and mechanisms borrowed from commercial enterprises.

Similarly, Trowler (1998a:94) argues that:

> hard managerialism elevates the power and role of management and the goals of economy, effectiveness and efficiency to a paramount position at the expense of many traditional stakeholders. Soft managerialism sees the application of 'improved' management practices as providing solutions in a difficult environment which will be to the benefit of all.

In the words of Deem (1998:53), "soft" managerialism recognises the need to overcome inefficiency and ineffectiveness through rational mechanisms, whilst, at the same time, emphasising the need for explicit collaboration and consensus. "Hard" managerialism believes efficiency gains can only be achieved through the imposition of reward and punishment systems, since employees are basically untrustworthy and incapable of self-reform.

Personally, I find Trow's distinction between "hard" and "soft" managerialism more illuminating than Exworthy and Halford's distinction between the analytical, the collective, and the individual dimension, important though the latter is. It seems that disagreement can persist even when two authors are both operating in the analytical dimension, both

discussing ideological underpinnings. As we shall see, a minority of authors believe that managerialist values and beliefs are quite appropriate for educational contexts because managerialism is seen as enhancing professionalism, whilst the majority think they are quite inappropriate in educational contexts because managerialism and professionalism represent diametrically opposed paradigms. One might hazard a guess that the 'harder' the managerialism, the greater the degree of friction.

As I said, a small minority of authors see little inherent conflict between managerialism and professionalism. Currie and Vidovich (2000:136) suggest many supporters of new managerialism would claim that "an ideology of privatisation does not necessarily have to be accompanied by an erosion of collegiality". For others, including Tomlinson (1998), management by professional managers / non-teaching experts might actually enhance collegiality, since it will enable practitioners to devote more time and energy to pedagogy, thereby improving the service they are able to offer clients, the primary goal of collegiality. Neither Currie and Vidovich (2000), nor Tomlinson (1998) distinguish between "hard" and "soft" managerialism, but it would seem likely that their conceptualisation of managerialism is "softer" rather than "harder".

By contrast, Randle and Brady's (1997a:231) conceptualisation of managerialism seems "harder' rather than "softer", when they write of two conflicting paradigms (professionalism and managerialism), each having different goals and values, key assumptions, and management ethos. Similarly, Simkins (2000) contrasts managerialism with what he calls bureau-professionalism and finds irreconcilable differences at every level. In bureau-professional systems, managers are "socialised" into a specific professional field and make decisions by combining bureaucratic rules with professional discretion and judgement in order to serve the needs of individual clients. In more managerialist systems, managers are "socialised directly into the values of 'management' as a generic discipline" and make decisions using specialist management techniques in order to achieve organisational objectives and outcomes. Whereas bureau-professionalism is

concerned with the well-being, needs, and rights of the client, managerialism is concerned with efficiency, organisational performance, and customer-orientation. The two systems do not merely use different techniques of management; they are based upon diametrically opposed paradigms:

> In their normative or ideological form they represent sets of values and ideals which provide *competing* discourses to justify and explain particular policy and management regimes. In doing this, they embody *contrasting* assumptions about power and legitimacy.
>
> Simkins (2000:321, italics added)

In bureau-professional systems, professionals have primacy, whereas in managerialist systems, managers have primacy.

Likewise, Flynn (1999:18) believes that "although we should be cautious about using grossly simplified concepts of 'management' and 'professions', there are *fundamentally important contradictions* between their values and practices ..." (italics added). The same sentiment is echoed even more stridently by Pollitt (1993:25), who declares that "the transfer, during the last decade or two, of managerialism from private sector corporations to welfare-state services represents the injection of an ideological 'foreign body' into a sector previously characterised by quite different traditions of thought".

Given this apparent incompatibility, there are those who would prefer managerialism to a previous system they perceive to be fatally flawed. Advocates of managerialism in the public sector point to the "the chaos of the old regime" with its uncontrolled irrationality, and proliferation of ineffective rules (Clarke and Newman, 1997:66-67). Those in education write of British schools plagued by "contrived collegiality" or the "adept use of micropolitical manipulation" (Brundrett, 1998:311), for whom managerialism might prove a saving grace.

In similar vein, Simkins (2000:329) notes how supporters of new managerialism in Further Education see it as "a modernising alternative, offering new opportunities for coherence and efficiency in college policies in contrast to the perpetual 'War of the Roses' which arose from departmentally-based 'feudal' college structures".

Nonetheless, wholehearted support for managerialism remains very rare in the literature, both empirical and non-empirical. Much more common are harsh criticisms of managerialism, such as those found in Flynn (1999), Clarke and Newman (1997) and Pollitt (1993). Managerialism, it is argued, lacks coherence and is riddled with internal contradictions. Pollitt (1993:111-113) notes one such contradiction in managerialism's promise of both greater delegation / local autonomy and more centralised control. Flynn (1999:32) notes another in the demand for both greater efficiency / doing more-for-less and greater consumer responsiveness / increased quality.

Managerialism is also criticised, severely by Pollitt (1993), somewhat less so by Clarke and Newman (1997), for failing to acknowledge fundamental differences between most public services and private sector businesses. These differences are particularly acute in terms of the market-competition, the supply-income relationship, and "the distinctive status of citizenship in public-service transactions" Pollitt (1993:130). In other words, public services are rarely allowed to fail – a school might close, but alternative educational provision would still be forthcoming; an increase in supply rarely means a corresponding increase in income – hospitals do not actually want to attract more patients, since this merely results in limited resources being spread ever more thinly; and notions of individual duty in the cause of the common good find no parallel in the self-centred world of individual consumerism. Managerialism is therefore seen by the majority of commentators as inappropriate to any public sector, and particularly so within the field of education.

Similar criticisms of managerialism are also found in numerous empirical case-studies, conducted at a variety of educational institutions around the world. Some of the issues are practical, relating to time-constraints and a narrowing of research focus, whilst others are more ideological, relating to values, quality, control, surveillance, and the nature of student-teacher relationships.

Currie and Vidovich (2000) interviewed 156 respondents from three Australian universities, and found that two thirds of them felt the demands to provide data on quality had gone too far, with the result that people spent more time trying to demonstrate excellence and less time actually achieving it. "Busy work" was detracting from the "main game". "Perhaps the major impact of accountability that may lessen the quality of education for students and the quality of research for the community is the *time* it takes away from more important tasks academics perform" (Currie and Vidovich, 2000:147-148, italics added).

Randle and Brady (1997b:133) make the same claim based on their own case study of an FE College in the UK. They argue that lecturers regarded "most" form-filling and paperwork as "fruitless and irksome", although according to their own figures only 54.4% of respondents said such quality assurance research was "not beneficial" compared with 28% who were undecided and 17.6 % who presumably did find it beneficial.

As well as taking up too much time, the demand for accountability was also seen as leading to a narrowing of research focus, away from what aroused curiosity and towards what had commercial application (Currie and Vidovich, 2000:148). In similar vein, Deem (1998) argues that managing research performance may actually inhibit intellectual curiosity and creativity. Combining issues of both time and quality, Trow (1994:17) asserts that when universities have external quality assessment imposed on them, they are likely to see "real quality" decline because they will be spending more time on bureaucratic reports, and on "fitting" their activities

into the external framework than on real creativity and the "pursuit of complex excellence".

In addition, many authors take issue with managerialism's apparent reduction of "complex excellence" to mechanistic number-crunching, and technicist performance indicators. Elliott and Crossley (1997:85) quote several FE lecturers lamenting the fact that quality is now reduced to the observable facets of student learning, and takes no account of what is unpredictable or hard to measure.

As can be seen from the works cited above, most case-study research concludes that people who would classify themselves as teachers and / or researchers, at primary, secondary and tertiary level, see managerialism as largely inappropriate and unwelcome. In the college Randle and Brady (1997b:130) studied, 80% of respondents to their survey thought that the recent organisational changes had not improved the quality of the service offered to students, and 95% thought the changes, as a whole, had not enhanced student learning. This leads Randle and Brady (1997b:135) to conclude that academic managers have a different value system to lecturers.

Similarly, in the FE college he studied, Elliott (1996:8) found evidence of "a pervasive market ideology, implemented by senior managers who seemed to embrace a managerialist culture, but questioned by many lecturers who held an alternative and competing democratic ideology underpinned by a commitment to a student-centred pedagogic culture". Most lecturers felt that the introduction of a "business efficiency model of education" was "undesirable", although they did recognise that the college needed to improve efficiency and save money in order to survive.

Likewise, Deem and Johnson (2002:71) conclude that the Vice-Chancellors and Pro-Vice-Chancellors in their study "have a shared community of practices and shared values and beliefs which are not comparable with those found in academic areas at faculty, school, or department level". In other words, faculty and even middle managers within a discipline had

quite different values and goals to more senior managers with institution-wide portfolios.

Not surprisingly, the picture is rather different amongst those who would characterise themselves as academic managers. Lumby and Tomlinson (2000) interviewed eight male principals of colleges in the north of England and all of them claimed to share, along with their faculty, the same primary commitment to their students. Of course, some of the respondents could have been deliberately disguising their views, but extracts quoted by Lumby and Tomlinson (2000) show remarkable candour in other areas - one principal admitted to replacing the most expensive staff in order to maintain curriculum breadth – so this seems unlikely.

What emerges most strongly from this study is that both senior managers and lecturers claim to have the best interests of the students at heart. They do not have fundamentally different value systems. Instead, they have different perceptions on how their shared values might best be realised. Such a position finds resonance in the work of Currie and Vidovich (2000:143), who note that many of their respondents said "they were not opposed to accountability per se, but to the form of the current requirements".

2.4.5 Shifting the balance of power – students, teachers and managers

Regardless of whether or not managerialism is considered appropriate to educational contexts, its emergence has had a profound effect on the power relationships between students, teachers and managers, particularly at tertiary level, where students have achieved the legal status of adults.

Now no attempt at analysing power relationships would be complete without at least a passing reference to the work of Michel Foucault. In his opinion, "Power is not something that is acquired, seized or shared, something one holds on to or allows to slip away" (Foucault, 1981:94). It is not possessed by individuals or groups; instead, it becomes apparent only when exercised, and it can be exercised in different ways, at different times,

by different people, depending on the context (Foucault, 1980:95-98). In this way, "power relationships are mobile, they can be modified, they are not fixed once and for all" (Foucault, 1994:285).

As we have seen, "hard managerialism elevates the power and role of management" (Trowler, 1998a:94). It affords managers the "right to manage" (Randle and Brady, 1997b:125). It is usually perceived by both its supporters and its detractors, as investing the current managers of today's educational institutions with far greater powers than those traditionally enjoyed by their predecessors in more collegial seats of learning (Meyer, 2002:534). Not everyone would agree that "new managerialism seems intent on tightening the 'control noose' around academics' necks even further" (Reed, 2002:159), but for sure, "managerialism has produced forms of university work organisation that increase the power of management and diminish the autonomy of professional academics" (Winter, Taylor and Sarros, 2000:282). In this way, managerialism is generally thought to have altered the balance of power between the manager and the teacher in favour of the manager, at least in theory, if not in practice.

Correspondingly, as was implied by the quotations from Foucault cited above, the power relationships between teachers and students, and between managers and students have also changed. For Elliott and Crossley (1997:84), managerialism has resulted in the marginalization of students, and a consequent reduction in their level of influence. They draw attention to "an apparent shift in focus away from the individual student, representing a pedagogical orientation, and towards business efficiency, representing a managerialist orientation". Individual students are viewed as a commodity, being important only insofar as they contribute to the overall efficiency of the institution.

By contrast, Randle and Brady (1997a and 1997b) see managerialism as increasing the power of students, in two ways. Firstly, both managers and teachers must appease students as never before, because the institution

needs their custom: "The lecturer / student relationship has traditionally involved notions of common enterprise and mutual responsibilities. The customer / supplier notion is beginning to cut across this relationship" (Randle and Brady, 1997a:235).

Secondly, teachers, especially, must appease students because management are now using them as instruments of control. Whereas Foucault (1997) writes of "hierarchical observation", in which superiors, quite literally, oversee the work of subordinates, Randle and Brady (1997a:235) write of "surrogate surveillance", in which managers use students to keep teachers in check: "Marketisation has re-constituted the student as 'customer' and encouraged surveillance of lecturers through quality systems and complaints procedures" (Randle and Brady, 1997a:238):

> Taken at face value a formal complaints procedure may seem like a reasonable and democratic mechanism, designed to protect the student ... . However, the complaints procedure as constituted represents a potentially destructive instrument that could undermine the authority of lecturers in the perception of the student.
>
> (Randle and Brady, 1997b:133)

Unfortunately, and in contrast to the study by Elliott and Crossley (1997), cited earlier, no informant quotations are offered in support of this assertion, so we are not sure how far it represents the authors' own opinion, and how far it is a view shared by the case study participants. Whatever the case, Randle and Brady are not alone in suspecting an ulterior motive behind the desire for consumer feedback. Clarke and Newman (1997:62) cite two articles showing how information technology can be used for "invisible" monitoring, and how customer surveys can be used to "discipline" employees.

Clearly, therefore, the emergence of managerialism in Higher Education is shifting the balance of power between students, teachers and managers, in a myriad of ways. Undoubtedly, the manager's star is rising, for better or for

worse. What this means for the fate of the teacher's and the student's star remains to be seen.

2.4.6 The relationship between managerialism and appraisal

So far we have discussed appraisal in terms of evaluation and development. We have seen that some authors feel these two elements can and should be combined, but the majority of the literature highlights the fundamental contradictions between them. The same can be said of the relationship between managerialism and professionalism – a minority of authors feel the two can be comfortably combined, whilst the majority find them mutually incompatible, at least at a conceptual level. Of course, some writers, such as Pollitt (1987), Walsh (1988) and Hutchinson (1997), simply equate managerialism with evaluative approaches to appraisal, and bureau-professionalism with developmental ones, but others, such as Fidler (1995) and Poster and Poster (1997), find this neat one-to-one correspondence something of an oversimplification.

Thus, for example, we find Pollitt (1987:94) distinguishing between performance assessment systems that are efficiency-driven, imposed from above, and imply a hierarchical structure, competitiveness and "the right to manage", and performance assessment systems that are self-driven, more oriented towards professional development, more egalitarian, and less individualistic.

The same two extremes of the continuum are highlighted by Walsh (1988:352) when he writes, "...[we can] distinguish two basic forms of appraisal; managerial, control-orientated appraisal, which is individually focused, judgemental and hierarchical; and participative appraisal, which is collectively focused, developmental and cooperative".

In the same way, Hutchinson (1997:158) argues that:

> Generally speaking, it is possible to identify two broad emphases in the forms and methods which appraisal programmes take: that which is directed towards extending

> and maintaining managerialist control and accountability, and that which aims at the professional development of staff.

In the former approach, productivity, economy and efficiency gains are paramount; there is a competitive system of rewards; faculty are expected to accept without debate "hierarchically defined criteria" concerning best practice; formal, written records are kept to show how each individual has contributed to the work of the organisation; training is offered only when staff appear "deficient" in the skills necessary to complete the prescribed tasks; and appraisal documents include observational checklists and are used for "surveillance and control" (Hutchinson, 1997:159).

In the latter approach, by contrast, emphasis is placed on shared responsibility and individual contexts, with a view to tackling areas for improvement collectively and continuously. Self-appraisal is given prominence and "Decisions on the focus and form the evaluation of practice is to take are negotiated and enacted collaboratively" (Hutchinson, 1997:159).

Not everyone, however, is content with the bi-polar distinction discussed above. Fidler (1995:97) mentions a tri-partite division, incorporating a developmental approach, an evaluative approach and a managerial approach. Here, the managerial approach is not simply equated with evaluation, but represents a position somewhere between development and evaluation, with a focus on reconciling the needs of both the individual and the institution.

Similarly, Poster and Poster (1997:154-5) outline a four-cornered matrix (replicated in figure 2.1 below) in which approaches to appraisal are labelled developmental, laissez-faire, managerial, or judgemental. Any organisation will locate its own appraisal system somewhere on this matrix, depending on the prevailing organisational climate and where its priorities lie.

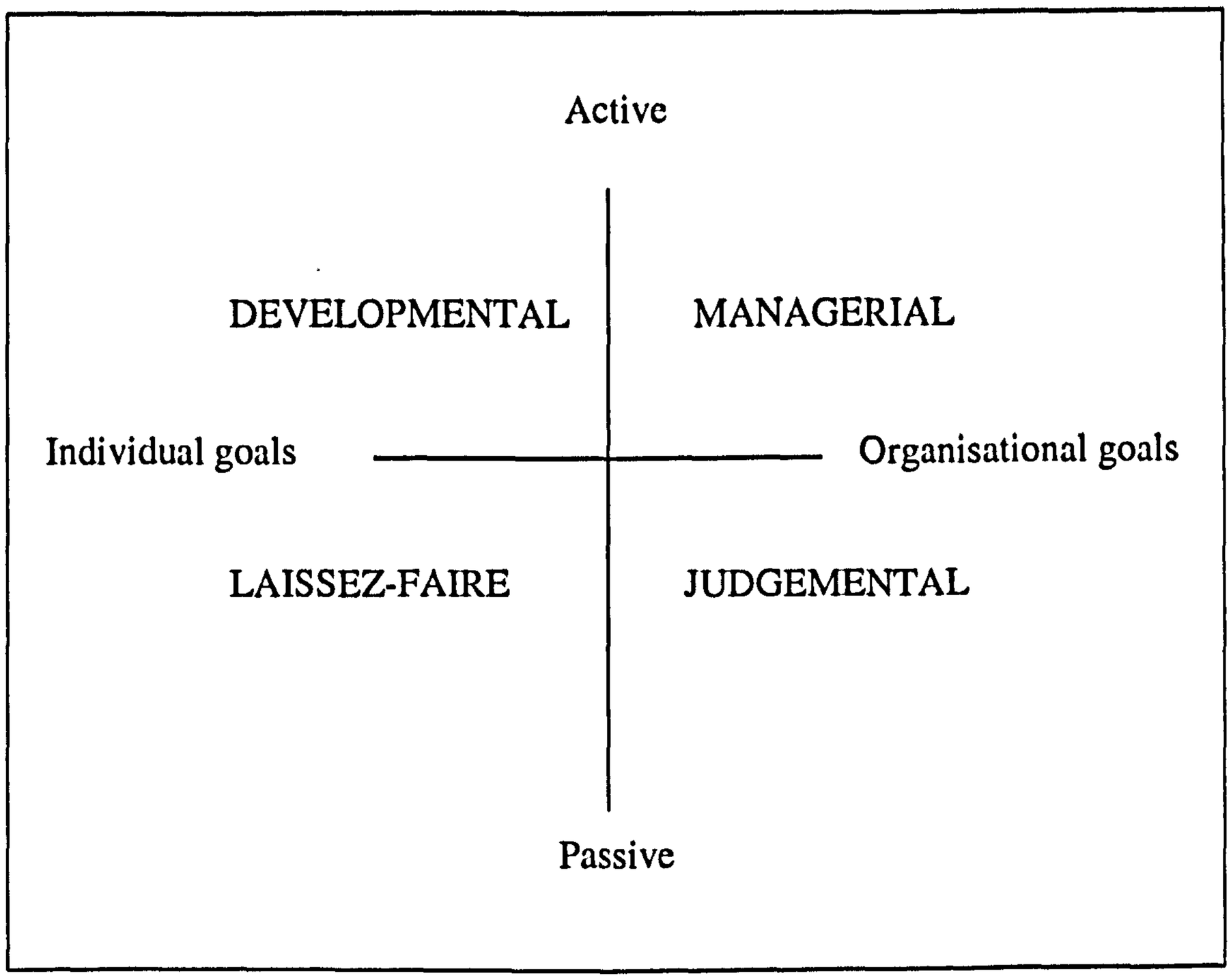

Figure 2.1

The managerial approach and the judgemental approach are similar in that both are oriented more towards organisational goals, whereas the developmental and laissez-faire approaches are both more oriented towards individual goals. The managerial approach is based on hierarchical authority and aims to maximise organisational efficiency and effectiveness through the setting of short-term performance targets and the use of incentives, praise, and reproach. The judgemental approach has a similar aim, but seeks to achieve this in a more authoritarian / controlling way, by rating individuals against each other and by assuming that the cultivation of extrinsic motivation (in the form of systematic performance related pay) will be a prerequisite.

By contrast, the developmental approach is based on collegiality and collective authority; it aims to uphold moral, ethical, and professional values through peer and self-appraisal; it tries to promote trust, openness

and co-operation, with an emphasis on self-directed, intrinsically motivated, longer-term personal and professional development.

The laissez-faire approach is similar to the developmental approach, but more emphasis is placed on the individual and less on the professional community. Neither the management nor one's peers drive the process – everything is left up to the individual, with a resulting lack of systematic focus, direction, and purpose.

In an area as sensitive as appraisal, within a debate as controversial as professionalism versus managerialism, it is easy to polarise extremes and caricature ideologies. It is tempting to set up a dichotomy between managerialism and bureau-professionalism, and then to match the former with evaluative approaches to appraisal, and the latter with developmental approaches. However, as we have seen, such dichotomies and one-to-one correspondences oversimplify a complex situation. Accordingly, I would suggest that Poster and Poster's (1997) matrix does better justice to the subtleties of the various positions, even though the friction between them may well be the same, regardless of whether it emanates from two sides or four corners.

Chapter 3: Methodology

3.1 Introduction

Blaxter et al. (1996:56) write of research families (quantitative and qualitative); research approaches (action research, case studies, experiments, and surveys); and research techniques (documentary analysis, interviews, observation, and questionnaires). According to this typology, my research family (though I prefer the term paradigm) is qualitative; my approach is case study; and my principal (data collection) technique is interview, with some documentary analysis and observation. Specifically, I conducted 38 semi-structured interviews at two different sites in the same Middle Eastern country between 1999 and 2002.

At Rihab University, I interviewed five appraisers / managers, along with fourteen appraisees (all EFL faculty members) over a nine-month period. Three of the appraisers were middle managers responsible for supervising a group of teachers; one was a more senior manager with overall responsibility for the department; and one was an academic advisor to the Provost with particular responsibility for faculty liaison. The fourteen appraisees were divided equally between those who, like me, had joined the university at its inauguration, in 1998, and those who had joined at the start of the second year.

Two years later, at Al Fanar College, I interviewed four appraisers / managers and fifteen appraisees (all EFL faculty members) over a four-month period. Again, three of these were middle managers responsible for supervising a group of teachers; and one was a more senior manager, with overall responsibility for the department. The fifteen appraisees included people who had varying lengths of service with the college, which at the time of the research was five years old.

In addition to these interviews, I undertook participant-observation, keeping detailed field notes for approximately two years at Rihab and one

year at Al Fanar, and analysed relevant documents relating to appraisal. I also kept a research diary in which I recorded what I was doing and reflected upon the research process and its emerging outcomes.

Undertaking research involves various activities such as identifying the focus of the study; deciding on the most appropriate research strategy or approach; using particular research techniques to collect data; interpreting the data thus collected in the light of previous research; and reconsidering the meaning of this previous research in the light of the new data. These various activities are sometimes seen as stages in a linear process, whereby the focus determines the strategy, which in turn determines the techniques, which in turn determine the results, which finally determine the conclusions. "Just as it is sometimes argued that your research questions should determine your approaches and techniques, so, in an analogous fashion, it is often suggested that the methods you use will significantly affect the answers you get" Blaxter et al. (1996:75).

Yet, to my mind, such a tidy sequence fails to capture the 'messy' reality of real-life research, whereby the researcher herself, her focus of study, her research strategy, and her methods of data collection and analysis all interact over time, to form the different strands of a multi-threaded helix. If I am honest, I chose to study faculty appraisal not so much because I thought it was a burning, yet under-researched, issue, as because it appealed to something in my personality, and related to something in my previous experience. Similarly, I chose to investigate the issue using a qualitative case study based on semi-structured interviews not so much because the logic of the situation demanded it, as because it seemed intuitively 'right'. Coffey and Atkinson (1996:146) acknowledge that sometimes "the logic of the research design and the strategies of data collection and analysis may well have been defined only after the project was substantially complete". This was true in my case but, as I hope to demonstrate, my findings are no less significant because of it.

3.2 A descriptive / exploratory case study: the right choice?

"A consensus has gradually emerged that the important challenge is to match appropriately the methods to empirical questions and issues, and not to universally advocate any single methodological approach for all problems" (Patton, 1999:1). Undoubtedly, this is true, but should not be taken to mean that the focus of the research is the only influence on the choice of method. The personality of the researcher and the logistics of the situation also play a part, and there seems little sense in denying this. Accordingly, I intend to outline the personal and practical factors that affected my choice of method, before going on to argue that, in spite of their influence, the methods I adopted remain highly appropriate, given my research focus.

If I am honest, I have a personal predisposition towards qualitative rather than quantitative research. Eight years ago, as part of a Master's programme, I did some quantitative research in the area of sociolinguistics, and, although my findings were well-received, I found the experience strangely mechanical and unsatisfying. By contrast, I did a qualitative case study as part of the requirements for a Master's in Education Management, and found the process much more exciting. No doubt this experience steered me towards choosing a research focus that *could* be located within a qualitative paradigm.

Similarly, my choice of a case study approach was partly influenced by practical considerations. I had a deadline for completing my research and my work commitments restricted my choice of site. The "bounded system" (Cresswell, 1996:36) of a case study, limited in both time and space, seemed to fit well with the constraints of a part-time researcher, hoping to submit a doctoral thesis in three years time. Thus, from both a personal and a practical standpoint, a case study approach seemed most suitable, and, by happy co-incidence, theoretical considerations pointed in a similar direction.

My research was concerned with "accessing the unique perceptions of

individuals" (Harrisson, 1994:42) and "reconstructing the social world of the participants" (Harrison, 1994:44). I was interested, not so much in uncovering some 'objective' representation of what faculty appraisal 'really is', as in discovering the range of complex, and perhaps contradictory, meanings that participants in this particular social drama gave (or at least claimed to give) to their experience of reality. Given this research focus, a grounded theory approach would seem to be most appropriate, since "Grounded theory methods are suitable for studying individual processes, interpersonal relations and the reciprocal effects between individuals and larger social processes" (Charmaz, 1995:28).

In addition, I began my research with a range of nebulous ideas about the significance and impact of faculty appraisal, rather than a clear hypothesis. Such a starting point would again suggest the use of grounded theory research where "rather than testing the relationships among variables, we want to discover relevant categories and relationships among them; to put together categories in new, rather than standard ways" (Strauss and Corbin, 1990:49).

I believed that my area of study was complex and quite sensitive, controversial even, and for this reason, I thought it better to concentrate on investigating one (or as it turned out, two) small and relatively self-contained communities, in considerable detail, rather than surveying a larger sample more superficially. Once again, this pointed to the use of a case study approach, which "allows the researcher to concentrate on a specific instance ... and to identify, or attempt to identify, the various interactive processes at work" (Bell, 1987:8), and in which "the researcher does not aim to cover a whole population and extract common factors, but to provide an in-depth picture of a particular area of the educational world" (Drever, 1995:7).

So, although my decision to adopt a case study approach was partly driven by practical considerations and no doubt influenced by my own personal predisposition towards qualitative paradigms, it also represented a sound, if

slightly inadvertent, choice, from a theoretical perspective. Similarly, whilst I did not initially decide to do a descriptive / exploratory case study, in retrospect, this best describes the approach I adopted, which seems none-the-worse for not having been pre-planned as such.

According to Yin (1993:5), there are three different types of case-study:

> An exploratory case study ... is aimed at defining the questions and hypotheses of a subsequent (not necessarily case) study ... A descriptive case study presents a complete description of a phenomenon within its context. An explanatory case study presents data bearing on cause-effect relationships – explaining which causes produced which effects.

Given this particuar typology, where "exploratory" seems to imply 'pre-cursory', I would say that my case study is more descriptive than exploratory. Similarly, given Bassey's (1999:63) typology, I would say that my case study is more "picture drawing" than "theory-seeking" or "theory-testing", because:

> picture drawing is predominantly a descriptive account, drawing together the results of exploration and analysis of the case ...[It] ... should give theoretical insights, expressed as a claim to knowledge, but this is more discursive than the fuzzy propositions and generalizations of theory-seeking and theory-testing case study.
>
> (Bassey 1999:63)

On the other hand, however, using Robson's (1993:42) typology, I would describe my case study as more exploratory than descriptive, since the former aims to investigate what is happening; pose questions; uncover fresh insights; and assess phenomena in a new light, whereas the latter merely aims to provide an accurate portrayal of people, events or situations. It is my intention to move from first-order to second-order constructs, from "descriptive to theoretical levels of analysis" (Strauss and Corbin,

1990:95), but still I would hesitate to say that my research is theory-generating, because of its very small-scale.

3.3 The criteria for judging research in a qualitative paradigm

As I mentioned earlier, my research rests firmly within the qualitative paradigm. Rather than trying to test a pre-determined hypothesis, it seeks to explore the issue of faculty appraisal in an open-ended way; rather than superficially surveying a large sample of people on one particular issue, it focuses on a small-scale case study, uncovering in considerable depth the perceptions of 38 people at two institutions; rather than using numerical data, categorised or rated at the point of collection, it works with 'unstructured' verbal data, recorded during semi-structured interviews.

This being so, the issues of validity and reliability become more complex than is normally the case in quantitative research. "The language of positivistic research is not congruent with or adequate to qualitative work" (Ely et al., 1991:95, cited in Cresswell, 1996:197). "Given that different epistemological and ontological assumptions inform qualitative and post-positivistic (quantitative) inquiry, it makes little sense to impose the criteria used to pass judgement on one upon the other" (Sparkes, 1995:185).

Yet, the precise form that qualitative criteria should take is still far from clear-cut. Backman (1981), Lincoln and Guba (1985), Hawkins (1990), Strauss and Corbin (1990), Robson (1993), Blaxter et al. (1996), Cresswell (1996), Boyatzis (1998), Eisner (1998), Bassey (1999), and Silverman (2000), all propose slightly different criteria by which one might judge qualitative research, some of them clearly owing more to the quantitative paradigm than others.

Thus, for example, Lincoln and Guba (1985:300) remain quite close to quantitative traditions when they suggest replacing internal validity, external validity, reliability and objectivity, respectively, with credibility (meaning whether or not the findings are 'true'), transferability (meaning

whether or not the findings are applicable to different settings), dependability (meaning whether or not the findings would be replicated if the study were repeated) and confirmability (meaning whether or not researcher bias has been minimised). Similarly, Bassey (1999:75) suggests that one should look for "trustworthiness" or "the ethic of respect for truth".

Other writers are less concerned with external 'truth', whatever that means, and more concerned with internal "consistency" (Boyatzis, 1998, and Silverman, 2000) or "coherence" (Eisner, 1998). Additionally, Eisner (1998) values "agreement amongst experts" in the same way that Robson (1993) prizes "intersubjective agreement" and Cresswell (1996) credits "peer review". Moreover, some writers, though by no means all, insist that qualitative research should have a practical impact. Thus, Blaxter et al. (1996) list "significance" as one of four criteria, while Eiser (1998) lists "instrumental utility" as one of three criteria. This is clearly related to the issue of how far the findings of a case study are generalisable to other settings, a key concern that will be looked at in the next section.

On a more practical level, Cresswell (1996:201) suggests that worthwhile qualitative research should include the following features: "prolonged engagement and persistent observation in the field"; triangulation of data; peer review; negative case analysis, in which disconfirming evidence is actively sought out; clarification of researcher bias; and informant validation. All of these features apply to my own research. Thus, for example, I was a participant-observer in each of my research settings for a period of years; I used interview data along with field notes and documentary analysis; I had my tutor and fellow doctoral students comment on my work; I systematically reviewed my data for both positive and negative elements, as will be demonstrated later; I kept a research diary detailing my reactions to the research process and, in writing up, have made no secret of the personal circumstances that may have biased my findings; and initially I undertook some informant validation, though this was discontinued for reasons that will be outlined below.

3.4 The generalisability of a case study approach

Obviously, as was mentioned before, the case study researcher is not trying to extract features shared by the whole population (Drever, 1995:7). "The case is studied in its own right, not as a sample from a population" (Robson, 1993:5). Generalising from a case study therefore means "extrapolating" themes that may prove useful in understanding similar cases, rather than merely demonstrating how far a particular instance is representative of a larger population (Silverman, 2000:111).

Stenhouse (1985:267) argues that comparing one's own case to someone else's case study allows fresh insights. Similarly, Stake (1995:69) suggests that case studies have interest for readers because they appeal to their "tacit knowledge", and that, if the experiences they read about are similar to their own, this can form the basis for "naturalistic generalisation". In this way, generalisability in qualitative research "is best thought of as a matter of the 'fit' between the situation studied and others to which one might be interested in applying the concepts and conclusions of that study" (Scofield, 1993:109). It is incumbent upon the writer to include as much thick description as possible so as to provide the reader with "the information necessary for an informed judgement about the issue of fit" (Scofield, 1993:109), but ultimately, "the problem of generalising ceases to be a problem for the author. It is the reader who has to ask, what is there in this study that I can apply to my own situation, and what clearly does not apply?" (Walker, 1980:167).

That said, I would still like to suggest to the reader three ways in which my case study might be a point of significant, rather than idle, comparison. Firstly, the fact that I investigated two very similar sites has allowed me to build up a more cumulative picture, with the result that my "fuzzy generalizations" (Bassey, 1999:62) are rather less tentative than they might have been, had I restricted myself to a single site.

Secondly, the appraisal systems at Rihab and Al Fanar appear to have much in common with systems elsewhere, in the West, but the "wider structural,

economic and political frameworks" (Grace, 1985:3) surrounding them are quite different, in ways that will be discussed in a subsequent chapter. The case study therefore offers one illustration of how appraisal might work in situations where employers face very few legal constraints and profess a concern for teaching excellence that appears to override all other considerations.

Thirdly, I believe that my reflections on my evolving research methodology have yielded insights that could be particularly valuable to other insider researchers. Case studies usually concern a single research site. Occasionally, they compare two sites, but this is normally from an outsider's perspective. I have not yet come across any other case study that centres on insider research at two separate but quite similar sites. For this reason, I feel justified in claiming that my methodological reflections and my comparative findings constitute an original, albeit rather small, contribution to the academic community.

3.5 The insider / outsider continuum

3.5.1 Definitions of terms

'Outsider' research is where the researcher is not *a priori* familiar with the setting and people she is researching. "Traditionally, the accepted approach to ethnography has been for complete outsiders to learn to be like 'natives', all the while retaining a distance and the perspective it provides" (Hockey, 1993:201). "Being a stranger, an outsider in the social setting, gives the researcher scope to stand back and abstract material from the research experience" (Burgess, 1984:23). By contrast, 'insider' research takes place when the researcher has *a priori* knowledge of the setting and / or is investigating her peers.

Certainly, outsider research has a longer history, and, in the eyes of some commentators, a more refined pedigree than insider research. Authors can still be found who advise against researching where one works (see, for

example Blaxter et al., 1996:29 and Cresswell, 1996:114) or else suggest that "strangers make the best informants" (Powney and Watts, 1987:50).

3.5.2 A continuum, not a dichotomy

Notwithstanding the above, the amount of insider research has grown steadily over the years, and, in any case, "the notion of a dichotomy between insider (administrator) and outsider (academic) research, although useful in calling attention to the unique problems entailed in each, tends to be somewhat overstated" (Anderson and Jones, 2000:443). Indeed, according to Hockey (1993:201) "the insider / outsider (or auto-ethnography / ethnography) dimension is best seen as a continuum rather than as a rigid dichotomy". In other words, one can have a varying degree of familiarity with the setting one is researching and the people one is interviewing. Or else, one can be an insider with respect to one part of a subculture, but not another. This has certainly been my own experience.

At both Rihab University and Al Fanar College, I undertook insider research, but there were very significant differences between the two processes. At Rihab, I engaged in particularly 'intimate' insider research in that I was well known to most of my informants for eighteen months before my research began, and had freely expressed my opinions on my research topic. By contrast, at Al Fanar, my research was of a much less 'intimate' nature because my informants had known me less than a year, and I had deliberately chosen not to discuss in any context anything related to my research topic. This difference had particular implications for my interviewing style, a point to which I will return later.

Somewhat surprisingly, according to Anderson and Jones (2000:433), "Many of the administrators who did studies within their own settings did not perceive their research as different from outsider research". By contrast, Hockey (1993:200) argues that whether one is an insider or outsider researcher may affect the whole research process, including site selection, sampling, documentary analysis, observation, and the way meaning is constructed from the field data. From my own experience, I would endorse

Hockey's argument and extend it by suggesting that, even within insider research, the various outcomes of the research process are significantly influenced by the degree of 'intimacy' that an insider researcher achieves at one site compared to another.

Clearly, conducting insider research is like wielding a double-edged sword. What insider researchers gain in terms of access and rapport may also be lost in terms of being known prior to the research (Hockey, 1993:206). What they gain in terms of "their extensive and intimate knowledge of the culture and taken-for-granted understandings of the actors" may be lost in terms of "their myopia and their inability to make the familiar strange" (Hawkins, 1990:417). The more 'intimate' the insider research is, the sharper the blades of the sword, and the more magnified the potential advantages and disadvantages become.

3.5.3 The disadvantages of insider research

In terms of specific disadvantages, being too familiar with the situation may mean one cannot see what would be obvious to an outsider. "This lack of 'the new', the relative absence of culture or entry shock (as experienced by the more traditional researcher in the role of outsider), may result in the researchers 'ignoring commonplace but meaningful behaviour and avoiding or subverting treatment of sensitive issues' (Jarvenpa, 1989)" (Hockey, 1993:206).

Moreover, in insider research, it is all to easy to take things for granted, to develop myopia and to assume one's own perspective is far more widespread than it actually is. An insider's familiarity may help them understand the social processes and meanings of a particular world, but it may also prevent them from challenging assumptions (Hockey, 1993:202). In this way:

> Confession can probably be made more freely to an insider, since shared group membership makes similar experiences more likely. For the interviewer, the availability of such experience, and of a shared 'public knowledge' tempts to

carelessness. Not to accept the point quickly, without requiring detailed explanations, can seem pedestrian and unintelligent, but may nonetheless be necessary to get clear and explicit data .... When it is assumed that norms are shared, their rationale and content do not need explanation, and thus the data become thinner.

(Platt, 1981:82)

A third disadvantage of insider research concerns the question of 'respondent bias'. Undoubtedly, "people's willingness to talk to you, and what people say to you, is influenced by who they think you are" (Drever, 1995:31). "Known or expected alignments or loyalties are crucial to the way in which an interviewer is perceived ... an additional problem for participant observers is where involvement in a community is likely to distort relationships with informants" (Powney and Watts, 1987:40). Though a certain degree of 'respondent bias' seems inevitable, in all social science research, the potential for distortion seems greater in insider research, because so much more is already known (or thought to be known) about the interviewer's stance.

A final disadvantage for insider research concerns the difficulty of addressing "obvious" (Hockey, 1993:206) questions to informants who believe the researcher already knows the answer because of their shared history in the same setting:

> If colleagues know that the interviewer is familiar with an event or experience which is the subject of the interview, they will be somewhat puzzled if the interviewer stimulates ignorance for research purposes. It would be too pedantic and stilted then for the interviewer to require a detailed explanation of an experience which is well-known to all participants through their common membership.
>
> (Powney and Watts, 1987:186)

3.5.4 The advantages of insider research

To my mind, the disadvantages of insider research outlined above are more than compensated for by an equally long list of advantages. Clearly, the insider researcher does not suffer from 'culture shock', but this need not be such a bad thing, since the culture shock experienced by an outsider researcher may actually have a negative impact, obscuring the researcher's vision and / or judgement (Hockey, 1993:203).

Moreover, access is more likely to be granted to the insider researcher and data collection may be less time-consuming. There is no traveling involved and more flexibility with regard to interview times. On the other hand, Scott (1985:120) contends that in participant-observation, it is often harder to tell where the research stops and the rest of life begins. I certainly felt this during my own data collection phase, which very quickly became 'all-consuming', since I was 'on-site' eight hours a day, five days a week, always on the look-out for incidents to include in my field notes.

Furthermore, some authors (such as Hockey, 1993) argue that insiders are more "invisible" and less intrusive, thereby altering the research process less, whereas others (such as Hawkins, 1990) suggest that a participant-observer who continues to perform their normal role within the institution will have more impact on the research than an outside consultant. Hawkins was the head of the school he was researching, so his suggestion is probably true of his own situation, but less so of mine, since I did not have any direct role in the appraisal system (other than being appraised as a faculty member).

Certainly, the insider-researcher is likely to have a better initial understanding of the social setting for a number of reasons:

> First, there is an intimate knowledge of the context; secondly, a knowledge of contextual features or events; thirdly, teacher-researchers are in a position to view both the obvious links between situations and events and also to understand the more subtle or diffuse links; and finally, they

are also in a position to assess the implications of following particular avenues of enquiry.

Griffiths (1985:211)

In addition:

> The [insider] researcher knows his / her environment well, knows by instinct what can be done and how far old friendships and favours can be pressed, just when and where to meet up for interviews, what the power structures and the moral mazes and subtexts of the company are and so what taboos to avoid, what shibboleths to mumble and bureaucrats to placate. They are familiar with the organisational culture, the routines and the scripts of the workplaces.
>
> Hannabus (2000:103)

Finally, and perhaps most importantly, the insider researcher usually has considerable credibility and rapport with the subjects of her study, a fact likely to engender a greater level of candour than would otherwise be the case. Practitioners might view an outsider researcher with some suspicion, imagining him to be an ivory-towered academic on day-release. The same cannot be said of an insider researcher:

> In effect, because the wider social structure classifies the researcher and informants in a similar or identical fashion, this creates greater confidence between the parties ... One of the results of this trust and exposure to the most intimate of details is that the insider researcher is able to appreciate the full complexity of the social world at hand. The result is a potentially accurate portrayal, rather than a simplistic caricature (Romano, 1968).
>
> (Hockey,1993:204-205)

3.5.5 Strategies to maximise the benefits of insider research

Obviously, it helps for the insider to stay as mainstream as possible, trying not to display too many out-group characteristics and remembering that all insider knowledge is only partial. It also helps to choose informants via some form of purposive sampling rather than merely relying on self-selecting (and like-minded) volunteers, a mistake I made at Rihab, but did not repeat at Al Fanar.

Similarly, the insider researcher needs to find some way "to make the familiar strange: to maintain enough distance so as to ensure that the analytical half of the insider / outsider coin operates effectively" (Hockey, 1993:208). Burgess (1984) suggests this might be achieved through continuous question-posing (both to oneself and to others); writing what one has observed in as much detail as possible; and regularly reviewing and cross-referencing one's findings. Other authors agree on the need to write as many field notes as possible, in as much detail as possible, as soon after the event as possible. Clearly, the significance of some events can only be seen in retrospect, and recollections several weeks after the fact have far less value. Accordingly, at both Rihab and Al Fanar, I kept two extensive diaries, one in which to write up field notes, and one in which to write memos about the research process and its emerging outcomes. I can still vividly remember racing up the steps from the cafeteria to my office, desperately trying not to forget a word of the hour-long lunchtime conversation I had just been a part of!

Interestingly, Evans (1995) turned her own experiences into fictional stories as a way of maintaining distance between herself and her research. Likewise, Hawkins (1990) chose to write his field notes in the third person in order to maintain a certain detachment from his data, a strategy I personally found a little too awkward. For similar reasons, Hockey (1993:210) recommends moving away from the research site when analysing and writing up the data. Again, this was not a practical possibility in my own case, but, all the same, I did find it impossible to work on my doctorate in my college office, even on those rare occasions when I had

nothing more pressing to do. And a two-week holiday, spent entirely alone, hundreds of miles away from the site of my research, was the only way I could free up the necessary 'brain space' to make sense of the reams of data analysis I had done previously.

Finally, in order not to appear hopelessly naïve, it may be necessary for the insider interviewer to acknowledge certain realities that are considered to be common knowledge by the informants, and to phrase their questions accordingly (Hellawell and Hancock, 2001: 4). Berreman (1987), for example, found he got a better response from people when his questions assumed polygamy existed, than when he tried to find out whether or not it existed. On the other hand, this strategy can obviously rebound if the meanings the interviewer assumes are shared, in actual fact, are not. In all honesty, and with the benefit of hindsight, I think there were definitely times at Rihab when I allowed a long shared history and an intimate involvement with my informants to inflate the amount of common knowledge I inadvertently thought could be assumed. I took care to avoid the same mistake at Al Fanar.

3.6 Research instruments: field notes, research diaries, documentary analysis, and semi-structured interviews

As I said earlier, for two years at Rihab and one year at Al Fanar, I wrote field notes of any incidents in which appraisal was discussed. These included formal meetings between managers and staff, which I minuted at the time, as well as lunchroom chats and passing remarks on the stairs, which I usually wrote up immediately afterwards.

I also kept a research diary in which I recorded my thoughts on the research process and my ideas about the emerging data. In addition, I carefully filed any documents relating to appraisal. Obviously these included examples of the various research instruments (lesson observation feedback forms, student evaluation questionnaires, summative evaluation forms, and so on). They also included a great variety of internal emails on the subject. These

were catalogued, but not routinely analysed in any great depth. Since my main research instruments were semi-structured interviews, I tended to analyse only those documents that appeared particularly relevant to something an interviewee had said.

Lastly, of course, I conducted 38 semi-structured interviews, with both appraisers and appraisees. I had intended to interview 20 people at Rihab, and then re-interview around six of them at various stages during the third year appraisal cycle. I had conducted nineteen interviews (five with managers and fourteen with faculty) when I received the entirely unexpected news that my contract was not going to be renewed. I immediately stopped interviewing, and although I continued to write field notes, I later discarded these, believing them to be irredeemably biased. Eight months later, I started a new job at a federal HE college, ten miles from the original site, and again began recording field notes and keeping a new diary. Over a four-month period in the second semester, I conducted another nineteen interviews (four with managers and fifteen with faculty), enabling me to compare perceptions from both sites.

3.7 Semi-structured interviews as the main data collection technique

3.7.1 Rationale

As I said earlier, I was interested in accessing opinions and beliefs on a sensitive topic in as much detail as possible. The issue of faculty appraisal I believed was complex, personal, and open to a very wide range of interpretation. Semi-structured interviews therefore seemed the most appropriate research technique since they "are especially suitable where one is particularly interested in complexity or process or where an issue is controversial or personal" (Smith, 1995:10). Moreover, the semi-structured interview "facilitates rapport / empathy, allows a greater flexibility of coverage and enables the interview to enter novel areas, and it tends to produce richer data" (Smith 1995:13).

I could also have used a questionnaire but I chose not to, because I felt that, no matter how well-written and extensively piloted it might be, it would still remain a fairly blunt instrument for uncovering perspectives about a subject as potentially sensitive as appraisal.

I am supported in my decision by Robson (1993:243), who contends that self-completed questionnaires are "necessarily superficial" because "there is little or no check on the honesty or seriousness of the responses. Responses have to be squeezed into predetermined boxes which may or may not be appropriate" and by Scott (1985:124) who contends that a survey approach hides and distorts the true complexity of a situation.

Furthermore, I noted that Cullen (1997), Gunter (1996 and 1999), Kyriacou (1995) and Hellawell (1991) all investigated the issue of appraisal through semi-structured interviews, without recourse to questionnaires, thereby producing articles that seemed to have greater "verisimilitude" (Hawkins, 1990:59) than those produced by Taylor (1997) and Rutherford (1988), who each relied on questionnaires. Indeed, Hellawell (1991:16) specifically argues that the issues surrounding appraisal are "sensitive and complex enough to suggest that questionnaire responses would be inadequate even if they were forthcoming".

Similarly, Currie and Vidovich (2000), Deem (1998), Elliott (1996), Elliott and Crossley (1997), Hellawell and Hancock (2001), Lumby and Tomlinson (2000), Shelley (1999) and Trowler (1998a) all investigated faculty perceptions of managerialism through interviewing alone, while only Randle and Brady (1997a and b) used both interviews and questionnaires.

Interestingly, the Committee on the Use of Human Subjects in Research at Rihab approved my application to interview faculty, without asking to see my interview schedule, but were quite insistent that I would need specific prior approval for any questionnaire I might wish to distribute. Although I was not intending to use a survey, this was just one more reason not to, as I

would not have wished my choice of questions to be constrained in any way by the university management. Similarly, at Al Fanar, management gave but then rescinded approval for a fellow researcher's questionnaire, a setback I was happy to avoid.

3.7.2 Sampling

"Sampling not only affects, but to a large degree, determines the degree of reliability and validity attainable" (Boyatzis, 1998:54), and is therefore a key consideration in research. In my own study, sampling was not an issue with respect to appraisers since I interviewed all of them at both sites. With regard to appraisees, however, I made the initial mistake of asking for volunteers at Rihab. This resulted in fourteen self-selected (and quite possibly like-minded) interviewees, most of whom happened to be female. To try and correct this bias, I deliberately conducted two further interviews with male faculty whom previous informants had recommended for inclusion in my sample.

I was careful not to repeat the same mistake at Al Fanar. Instead, I devised a system of purposive sampling whereby all of the English faculty were subdivided into four groups, according to the programme they taught and the manager they worked under. From each of these four groups, I decided to interview four people, ensuring, as far as possible, an equal balance with regard to gender and length of service. To this end, I had the Personnel Officer colour-code the faculty directory, highlighting those who, like me, were in their probationary year, and those who were 'old hands', having completed more than three years service at the institution. In two of the four groups, achieving the gender and length of service balance meant there was no room for manoeuve, and, luckily, all four of the people I had to approach agreed to be interviewed. In the other two groups, I began by approaching people in strict alphabetical order to try and avoid any bias on my part. This system broke down slightly on two occasions. In the first instance, a potential informant agreed to be interviewed, but declined to be recorded, and rather than proceed with the interview, I approached the next person on the list. In the second instance, I deliberately interviewed only

one of the three people who 'failed' probation, even though two of them appeared at the top of my initial list. I was able to interview about a third of the total number of English teachers and it therefore seemed more representative to include only one of the 'failed' probationers in the sample, as well as one person denied a contract renewal.

3.7.3 Place of the interviews

All but two of the faculty interviews were conducted in empty classrooms in order to ensure some privacy, away from the open-plan faculty areas. The other two faculty members were interviewed in their own private offices, and in retrospect I regret this because in both instances, the interviewees pulled documentation out of their filing cabinets and refered to it during the interview, something other faculty members were not able to do. This experience supports Smith's (1995:15) contention that the location of the interview has an impact on the data collected, and ought therefore to be as standardised as possible.

The management interviews at Rihab were conducted in their offices, which meant they were sometimes subject to interruptions. It also meant the status quo was maintained with regard to any preconceptions the interviewees and I might have had about the power-relationship between us. Having realised the drawbacks of my initial stategy, I decided to ask the managers at Al Fanar to move to an empty classroom I had set up in advance, an approach that seemed to work much better.

3.7.4 Timing of the interviews

Just as the place of the interview can affect the data collected, so too can the timing of the interview, often in ways the researcher cannot control. In order to minimise the effects of external events, I tried to conduct all my interviews as close together as possible, whilst still leaving enough time to reflect on (if not fully transcribe) one before beginning another. Unfortunately, this did not work out entirely as planned because part-time researchers, with heavy work commitments, trying to meet colleagues, with equally heavy commitments, do not have the luxury of planning their own

timetable.

So, at Rihab, the interviews were rather more spread out than I would have liked, with seventeen being completed during the five months from February to June 2000, and a further two being completed soon after the summer holidays. At Al Fanar, I was able to conduct all nineteen interviews within a four month period, from March to June 2002, though this meant that I sometimes conducted more than one a week, and once even conducted two on the same day, something I would never like to repeat.

Unfortunately, but unavoidably, some interviews were clearly influenced by a recent event that I had not been able to anticipate when fixing the time of the interview. For example, at Rihab, one interview was conducted on the day when staff were sent a memo informing them that, contrary to previous communication, they would be expected to work on a forthcoming public holiday. The interviewee referred to this several times and it obviously influenced some of her responses, particularly with regard to her view of management, a fact I noted but could do nothing about. Similarly, another interview took place on the day that the student evaluation results were given to faculty, and the initial part of our interview was dominated by this fact to such an extent that, in the end, I explicitly asked the interviewee to add her thoughts on other aspects of the appraisal system.

The issue of interview timing was even more critical at Al Fanar. After I had interviewed two people, it became clear (through the grapevine) that three probationers had been asked to leave, and this fact seemed to dominate the interviews conducted shortly afterwards. Unfortunately, I have no way of knowing whether the strength of feeling certain interviewees expressed at that time lessened as the semester continued. If I were seeking to uncover an 'objective, unchanging truth' this would be a problem. Since I am not, it is enough for me to be aware of these factors, and make sure I am not over-using (or indeed under-using) material gathered at this time.

3.7.5 Recording

All but one of the interviews were tape-recorded and transcribed verbatim. According to Smith (1995:18), the advantages of using a tape-recording so outweigh the disadvantages that he would never consider conducting a semi-structured interview without one. I entirely agree, and, that being so, when one of my potential faculty interviewees agreed to be interviewed but not recorded, I diplomatically substituted her with the next person on my list. Such a strategy was not possible with the appraisers, since I was aiming to interview all of them. Accordingly, when one of them asked not to be recorded, I went ahead with the interview and took notes instead. This particular interview took almost three hours and undoubtedly the data I collected were somewhat different in kind to the other data I collected via my recordings. I had to be selective in what I was able to write and when I returned my notes for validation, the interviewee made some substantial changes, deleting all of her expletives and toning down some of her more strident statements. I was unhappy about the extent of the changes she made but decided to accept them for the sake of consistency, since all the previous interviewees had been given the opportunity to validate their transcripts. Luckily, the situation did not arise again, although I subsequently discontinued the practices of respondent validation, for reasons outlined in a later section.

3.7.6 The interview schedule: how flexible is too flexible?

With regard to interview schedules, some authors (such as Brenner, 1981) set great store by sticking to the exact wording and order of each question in an attempt to achieve uniformity. I also think this approach ensures that all of the questions get asked, and facilitates very close 'question-by-question' data analysis. Still, it can feel constricting for both parties, and some authors, such as Smith (1995:15), acknowledge that a more interactive / conversational approach may yield more extensive data, accepting a certain amount of digression and reordering of questions in the interests of establishing rapport.

All the same, most authors still caution the interviewer very strongly

against revealing any of her own thoughts. Powney and Watts (1987:42) are typical of this stance when they write:

> By bringing in the interviewer's personal viewpoint, the respondent is not only distracted but may be in danger either of being acquiescent or of being prey to a self-fulfilling prophecy and giving the kind of information that coincides with what the interviewer apparently wants.

Similarly, Cresswell (1996:133) warns against sharing experiences with informants, since:

> This sharing minimises the 'bracketing' that is essential to construct the meaning of participants in phenomenology and reduces information shared by informants in case studies and ethnography.

Likewise, Platt (1981:77) writes of the need to resist the temptation:

> ... to contribute discreditable stories about oneself in anticipation and legitimation of return, to appear to get the point quickly without requiring explicit statements, and to treat the interview situation as one no different from other conversations and so contribute one's own quota of gossip and comment to the discussion.

On the other hand, Hawkins (1990:417) felt that the minimal responses used by ethnographers to elicit further information could be misinterpreted as a lack of interest. He also found that sometimes he had to give information in exchange for what he wanted from informants (Hawkins, 1990:416).

Taking this a stage further, Oakley (1981) argues that the interviewer ought not to withhold her own views or resist friendship and involvement because sharing experiences and attitudes helps to develop trust. Likewise, Griffin (1985) recounts how the young women in her study often turned her questions around by asking how she, herself, felt about leaving school and

getting her first job. "A positivist approach might treat these questions as irrelevant or unwanted intrusions, or as potential sources of 'data contamination' but I saw them as part of the reciprocal nature of the research process" (Griffin, 1985:102).

I am more inclined to agree with Hawkins (1990), Oakley (1981) and Griffin (1985) than with Powney and Watts (1987), Cresswell (1996) and Platt (1981), but all the same, I fear, that at Rihab, I may have been a little too 'chatty', particularly when discussing the student evaluation of teaching. Certainly, there were times when my interviewing style was overly 'interactive' with the result that I sometimes asked leading questions or provoked people into saying what they may not have meant.

After careful consultation with my tutor, I discounted any instances where I subsequently felt that I had put words into people's mouths. Having seen the error of my ways, I deliberately tried to be less conversational at Al Fanar and found to my pleasant surprise that, if I waited long enough and smiled encouragingly enough, people most often completed their sentences with the very phrases I had been itching to supply during the intervening pause. I also become more adept at deflecting personal questions during the interview, though I still allowed considerable digression from the interview schedule.

3.7.7 Respondent validation of transcripts and the status of interview data

As well as modifying my interview technique, I also altered my stance on respondent validation of interview transcripts.. At Rihab, I sent verbatim transcripts and summaries to all my informants and allowed them to make whatever changes they liked. Such a commitment to respondent validation is endorsed by Murray and Holmes (1997), and Robson (1993:34). It is also explicitly advocated by Bassey (1999:76), when he writes:

> It is good practice after the interview to take the reports of the interview back to the interviewee to check that it is an accurate record and that the interviewee is willing for it to be

> used in the research. Sometimes people realise that they have not said what they meant to say and this provides an opportunity to put the record straight.

Interestingly, as it turned out, very few respondents made any changes at all, perhaps lending support to Edwards and Furlong's (1985:27) contention that readers have neither the time nor the inclination to rewrite the account unless they are seriously misrepresented. A few people corrected typing errors, incomplete sentences or mishearings, but only two people made any kind of substantial change. The first instance, as previously mentioned, concerned the appraiser who toned down her strong language. The second instance concerned another appraiser who altered something she said because I had clearly been putting words in her mouth.

At Al Fanar, however, I decided not to allow respondent validation because I was won over by the arguments of Drever (1995:64), Anderson and Jones (2002:444), Silverman (2000) and Holliday, A. (personal communication, 17 March 2002) to the effect that respondent validation merely adds to your data without increasing their veracity. "Rather than commenting on the accuracy of your summary, they [the informants] are liable to want to expand and explain their answers, thereby introducing their own subjective bias into the interview record" (Drever, 1995:64).

Silverman (2000:35) raises the question of "whether interview responses are to be treated as giving direct access to 'experience' or as actively constructed 'narratives' involving activities which themselves demand analysis". He argues convincingly that informants' responses are not so much factual statements describing their experience of reality, as contextually-embedded narratives with a rhetorical force (Silverman, 2000:125). In other words, "Practitioners' accounts of their reality are themselves constructions of reality and not reality itself" (Anderson and Jones, 2000:44). They neither uncover some 'objective truth', nor reveal people's 'true' perspectives, firstly, because feelings, thoughts and emotions, by their very nature, are rarely static and clear-cut, and secondly,

because the researcher has no way of knowing how well the descriptions people give her correspond to their actual perceptions. The same person can have multiple understandings of reality, depending on the situation, and their verbal descriptions of these various understandings (be they 'genuine' or consciously contrived) will be different at different times and with different people. Accordingly, validation is "a flawed method" (Silverman, 2000:177) because it does not verify your data; it merely increases them.

3.7.8 Summaries versus 'cleaned up' transcripts

Initially, at Rihab, I based my data analysis on respondent validated summaries, taken from verbatim transcripts. These provided a useful shortcut to coding and could be read far more often than the transcripts, which were about five times as long. Naturally, I was meticulous in always going back to the original transcription and checking whether a code assigned on the basis of the summary was still appropriate. Nonetheless, as time went on, the use of summaries seemed less and less satisfactory, because of the inescapable degree of bias in the selection process (Drever, 1995:64).

Consequently, at Al Fanar, I worked only from 'cleaned up' transcripts, which had had some of their incomplete sentences, half-formed words and conversational fillers removed. To give just one example of what is meant by 'cleaned up' transcripts, one of my respondents was asked why she had said that we had low morale, and her exact response was as follows:

> "I wouldn't peg it, I mean, since this is an interview about the evaluation system, I wouldn't say that that's – I thin [*sic*] that that's symptomatic. The disease is really bad decision-making by the administration."

For the sake of clarity, this was rendered as:

> "... since this is an interview about the evaluation system ... I think that that's symptomatic. The disease is really bad decision-making by the administration."

My background in linguistics made me a little uneasy about tampering with the 'sanctity' of the spoken word in this way, but in the end I decided my aim in transcription should be maximum intelligibility, rather than complete accuracy. Powney and Watts (1987:147) contend that "any transcription is an *interpretation* by the transcriber of what is being said. What is written down is inevitably selective". This being the case, I eventually decided not to worry too much about omitting natural performance errors in order to make what was said more intelligible to a later reader.

3.8 Ethics

Powney and Watts (1987:147) argue that research benefits from interviewees being "fully informed from the start of what the researchers and the interviewees are trying to establish". For Silverman (2000:200), the issue is less clear-cut because researchers may "contaminate" their study "by informing subjects too specifically about the research questions to be studied". For Platt (1981:80), the dilemma is particularly acute when interviewing one's peers:

> Thus, it seems offensive not to give some honest and reasonably full account of the rationale and purpose of one's study to such respondents [who are equals] and the account cannot be one that is intellectually condescending. However, it is difficult to do this without inviting discussion of the study rather than getting on with the interview, and without providing so much information that it may bias the course of the interview.

After my experience at Rihab, I realised how easy it is to bias one's data collection, entirely inadvertently, and, therefore, I was deliberately much more vague at Al Fanar, telling people who asked only that I was studying an aspect of 'education management'. Indeed, this was the phrase I used when I approached people to request an interview, preferring to use the word 'appraisal' only at the start of the actual recording.

3.8.1 Field notes and overheard conversations

Griffiths (1985:210) describes how she chose not to use data from informal staffroom chats, or meetings with restricted access because the collection of these data had not been negotiated: "To release such data would be a betrayal of trust and an abuse of access. Herein probably lies another key to the research position, and that is the need for an understanding of the difference between research and voyeurism". Campbell (2002:41) felt somewhat the same, prefering to use only data from direct personal conversations, rather than anything he overheard by chance.

I, however, like Pollard (1985) and Scott (1985), decided not to adopt such an approach, partly because, unlike Campbell, I did not hold any position of responsibility in either institution I researched, and partly because I had no intention of making my findings available to anyone at my place of work. As it turned out, I used very few field notes directly, and none that involved overheard conversations.

3.8.2 Interviews: confidentiality and contradictions

I was in two minds about how much anonymity to give my informants. Obviously, Rihab and Al Fanar are pseudonyms. I could also have referred more specifically to my individual informants by allocating each of them a suitable pseudonym. If I had done this, the reader would have been able to see more clearly exactly who said what, and perhaps build up a better picture of each individual's perspective. On the other hand, this would also have enabled people from each institution to identify each other, since some of my informants, such as 'the head of department' or 'male supervisor', held a unique position.

A number of my informants made reference to the fact that they were telling me things in strictest confidence, and one of them even said she thought she would be "removed from the college immediately" if management heard what she had told me. For this reason, I thought it necessary to protect their identities as much as possible by being particularly vague. Accordingly, I have not attributed quotes to specific

people; I have distinguished between middle and senior management only where I thought this was particularly significant; and I have used the female pronoun throughout to disguise an informant's gender.

Another ethical dilemma centred on how to address contradictions in what people said. Sometimes someone said something in one part of the interview, and then seemed to contradict herself at a later stage; sometimes people said things that seemed contrary to my own experience; and sometimes they said things that were contrary to what other informants had told me. In the first case, I usually suggested I had misunderstood what they first said and asked for clarification. In the second case, at Rihab, I told people my own contrary experience, which I now realise was probably an unnecessary distraction. In the third case, I tried to suggest that other people thought or behaved differently, without being too specific.

This was particularly delicate when management at Al Fanar discussed the probationers who had been asked to leave. One of the managers had said, "it's our job to ensure that ... nothing is a surprise at any stage in the evaluation process at all". I had already interviewed one of the 'failed' probationers who did express considerable surprise at her dismissal, and several other interviewees who had similarly claimed that some, if not all, of the probationers had no idea they were not performing at the right standard. The following extract from the transcript shows how I tried, fairly unsuccessfully I fear, to challenge her assertion:

[Me] Coming back to something you said earlier about there should be no surprises in terms of the appraisal system; now, other people have expressed considerable surprise at some of the things that have happened. How do you square that with, in your opinion, there should not be any surprises?

[Manager] Can you tell me – can you give me a specific example that I can like ...

[Me] Okay, a number of people have been surprised at the non-passing – certain people did not pass probation and other of my interviewees have expressed surprise at that.

[Manager] Have they been surprised because they were appraised of all the facts through the entire process, or are they just on-lookers?

[Me] They're on-lookers.

[Manager] Well, that's because they've not been part of the process. The individuals who were involved in the situation that you've just described should have had it made clear to them that there were concerns from as close to the beginning as possible and those concerns should have been addressed and dealt with in the manner that I described to you earlier. And therefore the individuals concerned should not really have been surprised about the final outcome. Anybody outside of that loop may or may not be surprised – it's really – it's not a factor, I'm afraid, because, you know, it's personal and it's confidential, what went on between the [middle manager] or the [senior manager] and the individual.

In fact, as I said, one of my interviewees was more than just 'an on-looker', but had I revealed this to the manager, I would have been betraying her confidence, and I think the conversation may then have become too personalised, and the manager might have become distracted by trying to work out which of the 'failed' probationers I was referring to. So, although my challenge to her rhetoric was frustrated, I felt it more important to maintain confidentiality.

Slightly later in the interview, I feel I was more successful at drawing out an opinion that made reconciliation of these apparent contradictions

possible. I had steered the conversation around to another surprising appraisal decision, which I knew about outside the context of the confidential research interview. Someone in the same teaching team as me had been asked to leave and she had had some chats about it with various people (including me) in the staff-room. When asked about this situation, the manager said:

> ***[Manager] At the end of the day, the [middle managers] or the [senior managers] don't make the final decisions in these processes. Therefore, what I've described before can go on (pause) yet because the final decision isn't in the [middle manager's] or the [senior manager's] hands, (pause) that is, I think, as close to answering your question as I am gonna go.***

Obviously, in view of what I have already said about the status of interview data, I would hesitate before calling this extract more 'truthful' than the earlier one, but I do think it represents a deeper level of honesty, and illustrates how, very occasionally, an interviewer may be able to confront contradictions in such a way as to yield richer data.

3.9 Data analysis and grounded theory

Strauss and Corbin (1990:24) define the grounded theory approach as "a qualitative research method that uses a systematic set of procedures to develop an inductively derived grounded theory about a phenomenon". The researcher aims to build, rather than test, theory through an iterative three-part process in which "data collection, analysis and theory stand in reciprocal relationship with each other" (Strauss and Corbin, 1990:23). "This scanning and refining process, moving backwards and forwards between the raw evidence of the transcript and the developing analyses, is the classic constant comparison method of qualitative analysis first put forward by Glaser and Strauss (1967)" (Powney and Watts, 1987:105).

According to Strauss and Corbin (1990:23), the procedures followed during this process of constant comparison should be both "precise" and "rigourous". My own experience of grounded theory suggests that precision and rigour are laudable but well nigh unachievable aims! Data analysis is far more messy and subjective than Strauss and Corbin (1990) seem willing to admit. They do write of the need to balance science and creativity (Strauss and Corbin, 1990:45) but I think the phrases "chaos as a seed-bed for creativity" (Nias, 1991:162-3) and "moving from chaos to order, and from order to chaos, often simultaneously" (Blaxter et al., 1996:173) do better justice to the realities of grounded theory data analysis.

Similarly, I was struck by Maruyama and Demo's (1992:111) suggestion that once "all the instruments have been administered and returned ... we *just* need to take care of data coding, data entry, data analyses and data interpretation and we will be finished" (italics added). This, to me, gives an entirely false impression of what is an immensely complex and challenging process.

In my own case, this process was made even more complicated (but potentially more fruitful) by the fact that I conducted interviews at two different sites in two distinct phases. Strauss and Corbin (1990:205) suggest that:

> Once data collection begins, the initial interview or observational guides .... give way to concepts that emerge from the data. To adhere rigidly to initial guidelines throughout a study ... hinders discovery because it limits the amount and type of data that can be gathered.

Similarly, Charmaz (1995:34) argues that "data analysis drives subsequent data collection". Having to change my research site after two years felt very disruptive at the time, but, in retrospect, it provided an ideal opportunity for me to analyse the data collected at my first site very thoroughly, and then

devise a much more focused interview schedule for my second site, further details of which can be found in the following chapter.

I began my analysis of the Rihab data by rereading all of the transcripts and summaries over a period of two days. Gilham (2000:33) suggests reading no more than two transcripts a day, but I disagree because if one's reading becomes so spread out, one is more apt to forget what one read a week earlier. I then spent several days rereading the literature about managerialism and appraisal. With this background reading in mind, I finally embarked upon "open coding", described by Strauss and Corbin (1990:62) as "the naming and categorisation of phenomena through close examination of data".

I coded all the summaries on the same day, using very broad, mostly 'in vito' codes, and a code I labelled 'other' for instances where people said things which seemed significant but which did not yet fit into any other code. I sometimes assigned as many as four codes to the same line of summary, and I tended to assign more and more multiple codes as time went on because I began to see more and more connections between what various people said. Since Strauss and Corbin (1990:181) suggest "open" coding should aim to discover as many potentially relevant categories as possible, I was not at all perturbed to find that I had allocated 40 codes.

Sometimes I was coding the same opinion (even the use of the same word such as "perfunctory") but sometimes I was just grouping under the same code linked but contradictory themes. Sometimes two people would use the same word (e.g. "hoops") but the context was sufficiently different to suggest that they were not actually saying the same thing. And conversely, two people would use different words, but from the context they appeared to be saying the same thing.

After an overnight break, I recoded my transcripts blind, and was pleased to see there was a great deal of overlap between my first and second round of "open" coding. Where there was a discrepancy, I allowed the extract to be

coded in both ways. I then began "axial" coding, defined by Strauss and Corbin (1990:97) as " [putting] those data back together in new ways by making connections between a category and its subcategory". I decided, for example, that 'elements' and 'instruments' were in fact the same category; I also decided, for instance, that 'benefits' needed to be subdivided into 'management benefits', 'faculty benefits' and 'student benefits', and that many categories, such as 'faculty involvement', 'sense of shared ownership' and 'lack of time', needed to be cross-referenced.

I then used the cut and paste computer facility to build lists of the extracts relevant to each code. I went through these with a fine toothcomb, reorganising them, cross-referencing them, and meticulously refering back to the transcripts for each summary extract in order to be sure I was not misrepresenting what was said.

As an insider researcher, I was particularly concerned about avoiding perceptual bias, a pitfall I feel Strauss and Corbin somewhat underestimate. To be fair, they do recognise that:

> Each of us brings to the analysis of data our biases, assumptions, patterns of thinking, and knowledge gained from experience and reading. This can block our seeing what is significant in the data, or prevent us from moving from descriptive to theoretical levels of analysis.
>
> (Strauss and Corbin, 1990:95)

And they do emphasise the need to avoid presumption by constantly testing *a priori* literature-derived concepts and relationships against the specific data collected. But nowhere is there any statement as open as the acknowledgement by Miles and Huberman (1984:22) that "[while] ... 'Final' conclusions may not appear until data collection is over ... they have often been prefigured from the beginning, even when a researcher claims to have been proceeding 'inductively'". In similar vein, Eisner (1998:98) writes that "In fact, much of perception is at its inception interpretive".

Throughout the coding stage of my data analysis, I was conscious of the fact that I was engaging in a great deal of interpretation, and that, given my rather extreme circumstances, I might be finding only those negative elements I was subconsciously looking for.

As a corrective, I took heed of the advice suggested by Strauss and Corbin (1990:109), Eisner (1998: 110) and Blaxter et al. (1996:197), and specifically tried to seek out and then account for disconfirming evidence. Thus, for example, I systematically went back over all fourteen Rihab faculty responses to the question on the interview schedule, "Has your experience of appraisal had any effect on your relationship with the university management?", looking specifically for any positive notes. In this respect, I was following the example of Cresswell (1996:302), who found that "Because of the human cognitive bias towards confirmation (Mahoney, 1991), an active search for disconfirming evidence was essential to achieving rigour". I really do feel that I have genuinely looked for positive elements in the data, though, of course, without the benefit of a second researcher carrying out a parallel analysis, I can never be sure how successful I have been.

I engaged in a similar process of data analysis at Al Fanar, although, as I have already mentioned, this time I worked only from transcripts and not summaries. Because the interview schedule at Al Fanar was more focused, it was tempting to short-cut the coding process by simply relying on question-by-question analysis, but I decided against this because often I found people said things relevant to a particular interview question in a variety of different places.

I decided to code the Al Fanar transcripts 'blind' without reference to the codes I had devised for Rihab, because I was still concerned about how much my Rihab data analysis may have been negatively influenced by the non-renewal of my contract.

Having done "open" and "axial" coding on both the Rihab and Al Fanar

data, I began a final round of "selective" coding, described by Strauss and Corbin (1990:217) as "the integration of concepts around a core category and the filling in of categories that need further development and refinement". I reviewed all of the material from both sites, trying to adopt a more neutral stance with regard to the interpretation of the Rihab data, and extracting themes that would make sense of what was said in both places, and relate in some way to previous studies. The unique circumstances of federal employment in the country in question was a recurring theme and seemed to me to be the "core category" or "central phenomenon" (Strauss and Corbin, 1990:124) to which all other categories were in some way related. Once I had arrived at this conclusion, I set about finding some way of externally validating my own interpretation of the data.

3.10 Validation and triangulation of findings

Just as opinion is divided over the value of letting respondents alter their transcripts, so too is there controversy over the value of revealing one's findings to the reseach participants. For Edwards and Furlong (1985:33) the issue is clear-cut: "The major criterion of external validity is still the idea of presenting the researcher's account back to the researched. To be valid, an account must have convergence with the experience of the researched".

Unfortunately, such a stance fails to take account of the fact that the experience of an individual informant may alter over time, or else be ambivalent. In other words, feelings expressed during data collection may have been genuinely felt, at that time, but may have changed significantly in the intervening period before the data validation stage. This was particularly true at Al Fanar, where some unexpected decisions were taken with regard to probationary faculty, right in the middle of my interviewing cycle, no doubt altering some of the faculty's perspectives on the process. Validation in such circumstances would have been rather meaningless.

Similarly, some writers, such as Robson (1993), and Patton (1999), set great store by triangulation or "structural corroboration" (Eisner,

1998:110), whilst others, most notably Hodkinson (1998:17), Silverman (1993 and 2000) and Sofaer (1999), are less convinced. For Robson (1993:383):

> Triangulation in its various guises (for example using multiple methods, or obtaining information relevant to a topic or issue from several informants) is an indispensable tool in real world enquiry. It is particularly valuable in the analysis of qualitative data where the trustworthiness of the data is always a worry.

For Silverman (1993), quoted in Maggs-Rapport (2000:220-221), however:

> Using data to adjudicate between accounts forces the researcher to undercut one account with another. In effect, this ignores the context-bound and skilful character of social interaction. If accounts are context-bound, they cannot be verified by generating data in multiple ways and from multiple sources. Such data cannot be added together to produce a more complete picture, it is an end in itself.

Sofaer (1999:4) takes a middle line, implying that triangulation might be valuable, but can never yield the complete truth:

> Qualitative researchers often use the term 'triangulation' to describe the process of examining different perspectives in order to identify at a minimum what all informants seem to agree took place, or what it means ... This is almost as if we assume that the truth exists only in the space where multiple Venn diagrams converge. Perhaps it is more honest to admit that some of the truth may be found in the places in the diagram where the circles do not converge.

Accordingly, I have not reported my findings back to my participants (not

least because I thought this would affect how candid they wished to be), and I have used triangulation, somewhat sparingly. As I said earlier, interviews were my primary source of data, and I used official documents and field notes only in so far as they seemed to support or contradict my interpretation of what interviewees had said. Where possible, I drew out a consensus from what the majority of my informants had told me, trying to ensure that dissenting voices were still heard. At other times, where it seemed impossible to reconcile conflicting perspectives, I simply reported them.

I also took note of the five alternatives to validation and triangulation Silverman (2000:177) suggests one might use in order to increase the criticality of one's work. He recommends that, firstly, one should try to refute one's initial findings; secondly, one should constantly compare elements of the case study; thirdly, one should treat data comprehensively rather than selectively; fourthly, one should actively seek out and analyse deviant cases; finally one should tabulate or count one's data to ensure they are not misrepresented. All of these processes have featured in my own research. As I mentioned earlier, I deliberately sought out positive elements to try and counter the effects of my own negative personal experience; I was able to compare data from Rihab and Al Fanar throughout; and I have consistently used numbers to indicate how many of my respondents shared the same opinion. As a result, I feel justified in claiming that my findings have value, even though they were not subject to extensive informant validation or triangulation.

In addition, my findings have been further strengthened by extensive peer review, both from my doctoral supervisor, and from other doctoral students working in the same area. Obviously, these people are not in a position to gauge the accuracy of my raw data, but they can and have judged whether the data I have cited are sufficient to bear the weight of the conclusions I have drawn.

3.11 Endnote: transparency in the research process

I have endeavoured to describe my research process as accurately as possible, although I am aware that much published material does not display the same degree of detail and transparency. Most often, we are presented with "antiseptic accounts" (Measor and Woods, 1991:59) that give the impression of a seamless pre-planned progression from pilot study to final publication. I suspect that, in truth, this is rarely the case, unless the researcher is exceptionally skillful and / or exceptionally fortunate. Research, it seems to me, is an inherently complex, 'messy' process, and the research community, as a whole, would be better served by greater acknowledgement of this fact. Mistakes, of one kind or another, are inevitable, and should therefore be viewed as valuable learning opportunities, not only for those directly involved, but also for subsequent readers, since it is possible to learn as much from other people's mistakes as it is from one's own. For this reason, I am convinced that the research community stands to gain far more from the opportunity to read about what went wrong and evaluate how well it was rectified than it does from the pretence that nothing untoward or unexpected ever happens in our particular neck of the woods.

Chapter 4: Interview Schedules: The Generation of Questions and Answers

As has been mentioned earlier, I conducted my first round of faculty and staff interviews at Rihab between February and October, 2000. At this stage of the research, my intention was to investigate perceptions about faculty appraisal in a quite open-ended way. I was aware, from my preliminary reading, of potential links between the introduction of HE appraisal and the rise of managerialism. I was also interested in the management of change and how far participant involvement in policy formation was a pre-requisite of successful policy implementation (a focus I subsequently dropped). Accordingly, the interview schedule used with faculty at Rihab contained sections on the aims and objectives of the scheme; the extent of faculty involvement in the development and implementation of the scheme; the methods adopted; and its effects on management, faculty and students. The complete schedule is listed in Appendix 1.1.

The schedule for management at Rihab contained the same sections as the schedule for faculty, as well as two additional sections about the nature of appraisal feedback and the level of co-ordination between different appraisers within the same department. The complete schedule is listed in Appendix 1.2.

After carrying out further reading and preliminary analysis of the Rihab data, I made a variety of changes to the interview schedules. Following the advice of Smith (1995:15), I decided to begin each interview with a more general question about the interviewee's overall experience of appraisal.

I also decided to alter the section on aims and objectives slightly. When staff at Rihab were asked if the appraisal scheme was primarily evaluative or developmental, some of them gave ambiguous answers, or else said things that seemed to contradict assertions made elsewhere in the interview. So, at Al Fanar, instead of asking interviewees to make a bi-polar choice, I asked them to plot the college's position on an evaluative / developmental

continuum, thereby allowing them a better opportunity to expand on their perspective, and make more finely-grained distinctions.

I also asked them to comment on the assertion by Walsh (1988:357), written on a cue-card, that "the debate about appraisal is a debate about accountability, quality, competence, and professional autonomy". This enabled me to separate what interviewees thought about appraisal in general, from what they thought about the specific system in use at Al Fanar, and to investigate how far accountability models of appraisal were acceptable, in principle, to an academic community.

The section on faculty involvement was dropped completely, since the management of change was no longer a major area of interest, and this section had not generated data that were readily generalisable. Instead, I expanded the section on methods, to include a question about the appropriateness of the assessment criteria. This was done in order to explore how far appraisal was perceived by faculty as focusing on minimum standards of technical competency, as opposed to 'best practice' standards of creative adaptation, a key concern in the works of Peaker (1986), Turner and Clift (1988), Troman (1996), and Elliott and Crossley (1997).

I also added an entirely new section on time because this had been identified as a major issue in many similar studies of appraisal (such as Turner and Clift, 1988; Kyriacou, 1995: Cullen, 1997; and Randle and Brady, 1997a and 1997b). Appraisal was widely reported as very time-consuming and therefore I wanted to know how much time interviewees actually devoted to the appraisal process, and whether they thought this time was well-spent in view of the outcomes.

With regard to the section on appraisal effects, I realised that it was less confusing to ask six, rather than two, separate questions about management benefits, faculty benefits, student benefits, management drawbacks, faculty drawbacks, and student drawbacks. I also realised that, since "the order of

your questions affects what people have in mind when they answer each one, and this can influence what they say" (Drever, 1995:22), it would be a good idea to alternate the parts about benefits and drawbacks, beginning with benefits in some interviews, and drawbacks in others.

In addition, since three interviewees at Rihab had spontaneously mentioned the issue of grade inflation as a result of student evaluation of teaching, I decided to include a question about this on the Al Fanar faculty interview schedule. I also added a general question about the effect of the appraisal system on student-teacher relationships in order to discover, in what I hoped was a suitably subtle way, how far teachers felt under surveillance, as so strongly reported in the work of Randle and Brady (1997a and 1997b). Full details of the resulting changes can be found in Appendix Two.

It is a central tenet of grounded theory that previous data collection and analysis is used to inform subsequent data collection and analysis (Strauss and Corbin, 1990:205). This has most certainly been the case in my own study, whereby the data collected and analysed from my first site was used to inform the data collected and analysed from my second site. Accordingly, what started out as a very general and somewhat tentative enquiry into perceptions about faculty appraisal evolved, over a period of years, into a tightly-focused, in-depth investigation of the seven key research questions listed in my introduction and repeated in table 4.1 below.

1. *How is appraisal, in general, perceived by faculty and managers? How far are accountability models acceptable to either group?*

2. *Is appraisal perceived to have any effect on the quality of student learning?*

3. *What connections, if any, do teachers make between student evaluation of teachers and grade inflation?*

4. *How much time do faculty and management devote to appraisal? Do they consider that the time devoted to appraisal is time well-spent?*

5. *To what extent is appraisal said to focus on either basic technical competencies or flexible, creative practices?*

6. *How far is the specific appraisal system at each institution perceived as developmental and / or evaluative? How far is it seen as embodying a paradigm of professionalism and / or managerialism?*

7. *To what extent do teachers suggest they feel under surveillance?*

Table 4.1

Chapter 5: Appraisal at Rihab and Al Fanar

5.1 Contexts and Confidentiality

I suggested earlier that researchers should provide as much thick description as possible, so that readers could decide for themselves how far the results of any particular case study might be generalised to other contexts. Obviously, it is neither necessary nor desirable for researchers to include every single minute detail about their cases. They must make judgements about what is most significant, highlighting those features of the context they feel are potentially relevant to others, and leaving out those features they feel are completely unworthy of note.

Researchers must also protect the anonymity and confidentiality of their informants to whatever degree has been agreed between them. This is particularly important in cases where the focus of the research is deemed to be sensitive, or even controversial, in the eyes of either the researcher or the informants. Faculty appraisal seems to me to be an inherently controversial topic, and, as mentioned earlier, several of my informants spoke of the need to respect their confidence and protect their anonymity, one even going so far as to end our interview with the words, "very interesting discussion and I trust it will be anonymous".

Given this inevitable tension between offering the external reader thick description and promising the internal informant confidentiality, I would err on the side of caution. To me, it is more important to uphold the highest standards of confidentiality and anonymity, and thereby secure the continued support of research subjects willing to engage in candid discussion, than it is to provide the outside reader with intimate details. My aim, therefore, has been to try and ensure that my colleageues at Rihab and Al Fanar could not identify each other; and that external readers could not identify Rihab and Al Fanar. Accordingly, I have deliberately avoided naming the institutions and the country in which they are located.

Suffice it to say that Rihab and Al Fanar are both federally-funded higher education instutions offering a variety of qualifications to female undergraduates, entirely free-of-charge. The students are all citizens of the country in question and most of them have come straight from single-sex government schools, at the age of eighteen. The teaching staff are both male and female, in more-or-less equal proportions, and generally come from English-speaking countries such as Britain, Ireland, USA, Canada, South Africa, Australia and New Zealand.

5.2 Background to the appraisal system at Rihab

Rihab University was established in September 1998, in order to provide four and five-year Bachelor degree programmes in six disciplines, namely, Information Technology, Arts and Sciences, Communication and Media Studies, Business, Education and Family Sciences. During the first year, one teacher of English out of approximately 40 did not pass her one-year probation and left in the summer. During the second year, all the newly-appointed teachers passed their probation, and no-one else was asked to leave. As all the university contracts were for an initial period of three years, no contract renewal decisions were made during the time of the research.

Seventeen of my interviews were conducted during the second semester of the second year, whilst the final two were conducted very early in the third year. Most of my interviewees talked about the appraisal instruments used during the second year, but some who had joined the university at its inauguration also refered to the instruments used during the first year, which were not quite the same. For this reason, I have chosen to describe in some detail the appraisal instruments used in both the first and the second year of Rihab's existence.

5.2.1 The appraisal system at Rihab: Year one – 1998-1999

In the first year at Rihab, four appraisal instruments were used; namely, a peer evaluation form, a student evaluation of teaching (SET) questionnaire,

a lesson observation form, and a teaching and service dossier. Examples of these can be found in Appendix Three.

The peer evaluation form was handed out at the end of the first semester and each faculty member was asked to find a colleague to fill it out on their behalf. This was the only time the peer evaluation form was used. No official reasons was ever given for its discontinuance, but the manager responsible for its introduction said, in a later interview, that it had not worked as well as she had hoped, "because I think that people were very conscious of the fact that everyone was on probation and they were afraid to be candid about their colleagues". Indeed, one faculty member said she gave top marks across the board to the colleague she appraised and the colleague did the same for her. This was also my own experience of peer appraisal.

In the first year, the SET questionnaire was administered on paper to every class towards the end of each semester. Students were asked to tick boxes numbered one to four, according to how far they agreed with a variety of statements about their teacher and their course. The responses were anonymous and the form was completed in the presence of a teacher other than the one being evaluated. The results were taken in a sealed envelope to central administration and made available to faculty only after final grades had been turned in. Most faculty members taught two or three classes and therefore received two or three sets of results, together with a summary of all the scores.

Classroom observation by a manager was intended to happen every semester, but in reality, many people, including myself, had only one observation in the first year. The observer used a feedback form on which she recorded a score (one to five) for various aspects of the teacher's performance. The form also included a space for comments.

The teaching dossier was introduced, somewhat unexpectedly, in March of the first year. According to a subsequent memo from the Dean, the teaching

dossier was meant to serve "two primary purposes: providing the opportunity (and necessity) for you to reflect on your work thus far at [Rihab] and to help your [managers] better understand your goals, needs, and contributions". Faculty were expected to answer a series of questions about their philosophy of teaching, their view on the roles of students and teachers, their perceived strengths and weaknesses, and their goals for the future. They were also asked more practical questions about the classes they taught and the materials they produced.

In the first year, faculty did not receive any form of summative feedback, except a standardised letter in late May, congratulating them on passing probation.

5.2.2 The appraisal system at Rihab: Year two – 1999-2000

During the second year, several changes were made to the appraisal system. Examples of the revised instruments can be found in Appendix Four. As already mentioned, the peer evaluation form was dropped. The SET questionnaire was modified slightly and administered on-line. The lesson observation form categories were altered somewhat and the numbering system changed. Whereas in year one, the observer had to give a score out of five (with one being "needs improvement" and five being "does well"), in year two, the observer had to give a score of either one ("needs improvement") or two ("does well"). One of the managers I interviewed said this change had come about because she, and other colleagues, had found the five-point scale hard to use. She said:

> As an evaluator, I found it extremely difficult to work with one to five, especially with the wording as it was ... I remember the sentences to me just did not lend themselves to be able to split it up into a one, two, three, four, five category.

Similarly, another manager said it had been changed because "it was not possible to make finely graded judgements on the behaviours that were described".

Finally, the teaching dossiser was simplified and shortened. It was limited to two pages, and should include a list of classes taught, samples of materials prepared, a list of non-classroom contributions to the institution, and a list of scholarly or creative achievements.

In contrast to the first year, faculty, in May of the second year, received a more personalised three-paragraph letter from their Head of Department, listing some of the things they had done during the year and thanking them for their contribution.

5.3 Background to the appraisal system at Al Fanar: 2001-2002

Al Fanar was established in October 1997, as the latest addition to a nationwide system of vocational colleges, the first of which opened in September 1989. It offers a variety of undergraduate programmes, most of which are sub-degree level, lasting between two and four years.

As has been previously mentioned, in the year in which the research at Al Fanar was conducted, three people 'failed' their one-year probation, and were asked to leave at the end of the second semester, and at least one person who sought to have her three-year contract renewed was refused. In this respect, the situation at Al Fanar was quite different to that at Rihab, even though the appraisal system itself was quite similar, containing, as it did, the same three elements of student evaluation, lesson observation and self-evaluation.

At Al Fanar, students completed an on-line anonymous evaluation of their teacher towards the end of every course. This involved giving a numerical rating to their teacher in response to a number of questions (usually between twelve and twenty). Unlike the system at Rihab, there was no opportunity for students to write comments. Appendix 5.1 contains an example of the feedback that faculty received. Interestingly, at Rihab, faculty were quite deliberately handed the results of the student evaluations only after they had submitted their final course grades, whereas, at Al

Fanar, the results were made available to faculty as soon as they had been collated, usually before the end of the semester.

Faculty in their probationary year, and faculty who wished to have their three-year contract renewed had a formal lesson observation by their manager, twice in that year. Other faculty who were not on probation, and not up for contract-renewal were observed just once a year. The feedback forms used for this observation varied slightly from manager to manager, but all required the observer to rate various aspects of the lesson on a four or five-point scale. An example of a completed feedback form is given in Appendix 5.2

In addition, faculty were also asked to complete a self-evaluation form in two parts (see Appendix 5.3). Soon after the beginning of the year, they completed sections one to four, outlining their personal and professional goals for the coming year. They then met with their manager, ostensibly to discuss this document. Towards the end of the year, they completed sections five to seven, detailing their achievements during the year and indicating to what extent their goals had been met This process has no parallel at Rihab, where the teaching dossier was handed in only once, in April, without any formal discussion.

At the very end of the academic year, faculty received a quite detailed summative performance evaluation from their manager (see Appendix 5.4) including information about their student evaluations, their lesson observation, and their overall contribution to the college, as well as suggestions for goals in the following year. Again, such a document had no real parallel at Rihab.

Chapter 6: Findings

In chapter four, I included a series of more detailed questions that relate to my research hypothesis that much can be learnt about the appropriateness of managerialism in HE from studying people's perceptions of appraisal systems.

These questions came about partly as a result of rereading much of the literature on appraisal and managerialism, and partly as a result of doing preliminary analysis of the initial study. I have chosen to present the findings as 'answers' to these questions, although some answers, I feel, are more tentative than others. Answers two and three (about changes in classroom practice and grade inflation) seem to me to be the most substantive because the questions are quite narrowly focused and because question three was specifically asked of all respondents at Al Fanar. By contrast, answer one (about accountability models, in general) and answer six (about professional and managerial paradigms) seem to be the least substantive because these questions are much broader and much more complex than those relating to classroom behaviour and grade inflation. Consequently, I feel these findings are much more provisional at this stage, and will probably always remain very "fuzzy" (Bassey, 1999:62), however much further analysis is done. The other answers lie somewhere between these two extremes.

In reporting my findings, I have tried, as far as possible, to use numbers rather than more vague terms such as 'some', 'several', 'many' or 'most'. In interpreting these numbers, it is necessary to remember that between February and October 2000, I conducted nineteen interviews at Rihab, five with managers involved in appraising teachers, and fourteen with faculty members. Of the five appraisers, three were middle managers (the direct line-managers of the appraisees), and two were more senior managers. Of the fourteen appraisees, seven were faculty in their first year, and seven with faculty in their second year.

At Al Fanar, I also conducted nineteen interviews, between March and June 2002. Fifteen interviews were with faculty members being appraised; and four were with management doing the appraisal. Six of the fifteen faculty members were in their probationary year, and one of them had already been told that she had 'failed' probation and would have to leave at the end of the semester. The other seven who were no longer on probation included one person who was voluntarily leaving at the end of the semester and one person who had been told her contract was not being renewed after six years.

With regard to the four management interviews, three were with middle managers carrying out appraisals on the faculty members they line-managed; and one was with a more senior manager, who was not directly involved in the appraisal process, except in cases where managers recommended that a faculty member did not pass probation or did not have her three-year contract renewed.

In order to preserve anonymity as much as possible, each informant will be referred to as 'she', although, of course, many of them were in fact male.

6.1 How is appraisal, in general, perceived by faculty and managers? How far are accountability models acceptable to either group?

6.1.1 Findings from Rihab

Interviewees at Rihab were not asked any questions about appraisal, in general, as opposed to the specific system being implemented at their particular institution, a rather unfortunate oversight that was corrected on the Al Fanar interview schedule. Nonetheless, several interviewees did talk more generally about appraisal, either comparing Rihab with other educational institutions, or else comparing academia with business. All of those who talked in this way suggested that, in principle, appraisal was an acceptable facet of academic life, though two were keen to stress that business models were not applicable to educational institutions.

One interviewee told an anecdote of how, in the past, she and her fellow university students had used the student evaluation of teaching questionnaire to address what they saw as sexist behaviour on the part of a newly-arrived Brazilian teacher. She concluded the story by saying:

> I think, in the big picture, it [appraisal] is a good idea ... As a teacher, I'd rather not go through it. It's kind of like going to the dentist. But, it probably is good, overall.

Similarly, another interviewee suggested that appraisal was "a necessary evil" since "without appraisal, when you've got a large staff, you really don't know what people are doing because you can't be everywhere ...".

A third interviewee contended that appraisal was "necessary – for any institution to function, it has to have some system of evaluation". A fourth interviewee went into more detail, pointing out that:

> People get evaluated in their jobs all the time ... I'm not a business person at all. I fought it my whole life. But we're all accountable for our job, for what we do ... in every other job on earth, if you don't do your job well, you probably don't keep the same job.

These sentiments from faculty were also endorsed by one of the managers who suggested that, "[it] is a general requirement everywhere now that we evaluate the way that everybody is working ... to try to make them as effective as possible", a requirement she did not appear to consider onerous.

That said, two interviewees did highlight the dangers of importing business models of accountability into education. The faculty member who described appraisal as "a necessary evil" added that, although appraisal had originally come from business, sales people and teachers were completely different:

> You can say to a salesman, I want you to sell ten more cars this month. It's impossible to say to a teacher, I want you to

> pass ten more students this month, because you cannot control learning ... You can persuade people to buy something, but when they are learning a different language, and a different culture, and everything that that gives to people, it's just not the same.

One of the managers who had previously worked for many years in business agreed, saying:

> I don't think the two [business and academia] should be married. I really don't. I think they're two separate ways of running an organisation. You have two completely different mind-sets - a business person is intent on a particular goal for a large organisation and teachers have more finite goals that they're working towards in the classroom ... I'm not sure that it's a good idea to import a system entirely. There might be some things that you could learn from business, but giving value for money? Oh, I would hate to put it in those narrow terms.

Thus, it would seem that interviewees at Rihab were not opposed to the principle of appraisal, per se, as long as the inherent differences between business and education were taken into consideration.

6.1.2 Findings from Al Fanar

Having realised the need to separate perceptions about apprasial in principle from opinons about the specific system in place at a particular instituiton, I modified the interview schedule at Al Fanar accordingly. This time, I began by asking each interviewee a very general question about her experiences of appraisal to date, and a little later, asked for her reaction to the statement by Walsh (1988:357) that "the debate about appraisal is a debate about accountability, quality, competence, and professional autonomy".

Without exception, all Al Fanar interviewees agreed that an educational institution has the right to appraise the teacher's performance to determine competence and quality. One manager said she could not imagine how an organisation could possibly maintain any kind of quality without some kind of appraisal system, whilst another thought appraisal was necessary "if we are to regard ourselves as a profession". Similarly, one faculty member asked, "Why shouldn't an institution have an appraisal system to ensure quality among teachers? Why shouldn't it?" whilst another suggested, "there has to be some way of knowing if a teacher is completely wacko". A third faculty member referred to a previous job at a private language school without any formal appraisal, saying:

> I suppose I feel quite strongly about the fact that if the management becomes run down, and it's left up to just individual teachers, it just doesn't happen and it's dropped. And, although we may not like it, it's better for the school, the college and the teachers concerned that those measures are put in place, because otherwise, well, I thought I started to become unprofessional, because it didn't matter.

The general sentiments expressed by many interviewees were unwittingly summed up by the one faculty member who said, "most people I know don't have a problem with whether there should be an appraisal or not; it's about how it is conducted".

Indeed, whilst everyone concurred on the need for appraisal, they disagreed on the best way to judge quality; the extent to which teachers should be held accountable to different parties; and the amount of professional autonomy they should be afforded. With regard to quality, three faculty members drew attention to the same perceived difference between business and academia that had been previously mentioned by the two interviewees from Rihab. One of the Al Fanar faculty members said, "We are not a factory producing - where you can do your nice little production management studies, and efficiency schedules and time surveys. You can't quite do that. Teaching doesn't work like that". The other two drew

distinctions between companies "selling widgets" or "making cars" and "an academic institution" in particular, or "a service industry" more generally.

One of these two teachers then went on to insist that pass rates were a poor indicator of teacher quality, agreeing with Turner and Clift's (1988:98-99) contention that "Examination results are a measure of a long-term process and ... a very crude reflection of an individual teacher's input". By contrast, a different faculty member said that teacher appraisal should be based on test scores, as this was more "objective" and less "political" than other methods, such as the manager's summative report.

Most interviewees focused quite narrowly on the quality of the teacher. However, one faculty member did suggest that such a narrow focus, though very prevalent, was actually unhelpful. In her eyes, appraisal:

> should be about the quality of everything ... the quality of the materials you have ... the quality of the managerial support you have ... the quality of the appraisal even. But, it doesn't seem to be about any of that. It often comes down to the quality of the teaching and the quality of the teacher.

Exactly the same point is made by Walsh (1988:365). He suggests that any appraisal of an individual teacher should take into account the context in which she is working. He also suggests, however, that very often this is not done, precisely because the appraiser is partly responsible for creating the working environment out of which inadequacies may have arisen, and would rather blame the teacher concerned than admit partial culpability.

In the same way, some faculty members seemed comfortable with making teachers accountable to the institution, whilst others said teachers should be first and foremost accountable to the particular students they taught. One of the managers made explicit reference to this dual interpretation of accountability, speaking of a "personal accountability to students" that "most teachers feel everyday of their lives", and contrasting this with an accountability to the values, goals and rules of the institution. She

highlighted how these two kinds of accountability often conflict, particularly in relation to institutional rules on things like attendance and exam resits:

> It's kind of the personal and the institution. You know, you've got the two at war in a sense, when you're looking at appraisal.

Opinion was also divided on the question of how much professional autonomy was desirable; how much professional autonomy teachers at Al Fanar actually had; and whether any perceived curtailment of professional autonomy was part of a global rise in accountability, or merely the result of working in a particular Arab-Islamic context. Several faculty and three managers spoke of the need to trust teachers as competent professionals. However, two faculty members mentioned how the notion of professional autonomy could be misused by teachers. One said it was trumpeted by "people who don't like to be appraised [even though] you should open yourself up to other opinions". The other remarked on how "they call it professional autonomy, and what they really mean is 'I want to do exactly what I want to do. And you have no right to tell me what to do'".

Likewise, the senior manager suggested "too much professional autonomy leads to a whittling down of standards" particularly in relation to exam pass rates.

One faculty member and one manager mentioned the considerable freedom they thought teachers at Al Fanar enjoyed in terms of curriculum delivery, though not curriculum content. By contrast, three faculty members said they had less freedom at Al Fanar than they had had at previous institutions, but only one of them attributed this to the need for greater institutional accountability.

To sum up, it seems that neither the faculty nor the management of Al Fanar were opposed to appraisal in principle. Nor were the perceptions of faculty and management in obvious conflict. Undoubtedly, opinions

differed considerably on how to measure quality, and how to balance accountability and professional autonomy, but these differences were not split along partisan lines.

6.1.3 Comparison with the literature

The findings outlined above only partly concur with previous research. In studies conducted in the public school sector, almost all teachers are reported to have accepted the principle of appraisal. According to the work of Fitzgerald (2001:12), cited earlier, 98% of respondents agree that some form of teacher appraisal is essential to raise standards of teaching and learning. Likewise, 80% of the people in Middlewood's (2001:131-132) study saw appraisal as "essential for teachers' accountability" leading Middlewood to conclude that "the case for some form of assessment of teachers is accepted by the profession".

The position in Higher Education, however, is less clear-cut. Many studies in this sector point, not to the widespread acceptance of appraisal on the part of teachers, but to deep-seated divisions between faculty and management over the issue of professional accountability. According to Randle and Brady (1997a: 232), "85% of respondents believed that the college management did not share the same educational values as staff". Elsewhere, they write of "a conflict of paradigms" (Randle and Brady, 1997a:237) and "the emergence of a new type of manager in FE operating with an apparently different value system from that of the academic staff" (Randle and Brady, 1997b:135).

Similarly, the case study by Elliott and Crossley (1997:89) is said to have highlighted "a fundamental difference between lecturers and senior managers over the definition of quality, value, and improvement". The same point is made, albeit less stridently, by Deem (2000:15), whose own case studies threw up "some sharp contrasts between more optimistic stories of achievement and change told by manager-academics, especially at senior levels, and the more pessimistic accounts given by some support staff, Students Union sabbaticals, and ordinary academics". Such a sharp

dichotomy between the educational values of managers and faculty finds very little resonance in my own research. On the contrary, the range of opinion expressed by managers and faculty is almost as wide within each group as it is between each group.

6.2 Is appraisal perceived to have any effect on the quality of student learning?

6.2.1 Findings from Rihab

All fourteen faculty interviewees were asked about possible benefits of the appraisal system for the students, and whether the appraisal system had had any effect on their classroom behaviour. Four of the five managers also offered an opinion on the subject, even though they were not specifically asked about it. Two managers said the faculty were very capable, professional, and experienced, and, as a consequence, the purpose of the appraisal system was just to confirm their abilities rather than suggest improvements. By contrast, two other managers both suggested the appraisal system should lead to improvements in student learning. One of them was reported by faculty to conduct more than the one observation per semester prescribed in the official documentation and seemed to view observation as a way to improve pedagogy. In her words, the purpose of observation was "to see methodology in action and classroom interaction, with a view to identifying features of our approach which could be generalised and areas which we may need to refine and develop".

The other manager suggested that improvement might come, not so much from observation, as from student feedback, because "hopefully, it (the SET questionnaire) would point out at least some obvious areas that might need improvement, and that would benefit future students".

With regard to the faculty, one of the teachers reported having a very thorough (and helpful) post-observation debriefing, and said that aspects of her classroom behaviour had changed as a consequence. The other thirteen

teachers said their classroom behaviour had not changed as a result of the classroom visit, although four of them went on to say they wished they had received more detailed post-observation feedback, and would have welcomed the chance to try out any suggestions for improvement.

Two teachers reported changing their classroom behaviour as a result of feedback from the formal SET questionnaire. One had "changed a little bit" and the other had "made some changes". A third teacher said she changed "slightly" as a result of informal on-going feedback she had solicited from students.

So a total of four teachers at Rihab said they had tried to improve their teaching as a result of some aspect of the appraisal system. A further seven teachers said the appraisal system had not had any effect on their classroom behaviour, the strongest sentiment coming from someone who said, "I've learnt nothing about my teaching from the whole process ... so, in itself, it's useless for me". Interestingly, the remaining three teachers suggested that the SET questionnaire had actually had a negative rather than a positive effect, in that they now concentrated on keeping the students happy rather than helping them learn. One of these teachers, for example, recounted how she was "more focused on trying to please the students rather than trying to teach them English", making sure she did more computer work, more library visits, more videos, and fewer grammar exercises, even though this was not what she thought they really needed. These three teachers also admitted inflating their grades in a effort to boost their student evaluation scores, something that will be discussed in more detail in section 6.3.

6.2.2 Findings from Al Fanar

The fifteen faculty members were all asked, firstly, if the appraisal system had had any effect on their classroom behaviour, and, secondly, whether it had had any effect on their relationship with students more generally. At a later point in the interview, they were also asked whether the appraisal

system had any benefits for the students. This last question was also asked of the four managers.

Ten faculty members said that the appraisal system had had no effect on their classroom behaviour, although one of these suggested the detailed observation feedback she had received after one particular class visit, several years ago, may have improved her classroom teaching at that time. Of the remaining faculty members, three claimed it had had a negative effect on their teaching. One mentioned that all the non-teaching commitments she felt compelled to undertake paradoxically left her with less time and energy to devote to her students:

> I think that it [the appraisal system] affects how prepared I am for lessons, very often. And the fact that I often go to class tired and stressed, I think, it translates into the way I react towards my students. I don't maybe put the extra energy into developing a rapport with them, maybe, the way I should.

Another reported that she had become much less engaged with her students and much less inclined to challenge them to work harder because she feared what would happen if they complained. A third, who had not passed probation, spoke of teaching in a more natural way, once this decision had been finalised, implying that while this decision was still under review, her teaching had been somewhat strained.

By contrast, one teacher explained how her teaching had changed for the better as a result of feedback from the student evaluation of teaching:

> In the past, yes it has. I did notice consistently, that I ranked lower on [the SET category] "classes are interesting", and now, I really do make an effort to just move things around, juggle things around, break things up – so I do take those things on.

A second teacher said her teaching had also improved, though not directly through observation or student feedback. Rather, she had identified on her self-evaluation form a particular training course she wanted to run, and having been allowed to offer this course, she said that her own classroom teaching had improved as a result.

The evidence from this study therefore suggests that appraisal is not usually perceived as having any effect on a teacher's classroom behaviour, and even in those rare cases where a change is discernible, it is not always for the better. However, it is still possible that appraisal could be seen as having a positive impact on learning, not so much by improving the already adequate practices of most teachers, as by targeting the shortcomings of the failing minority. This was a view endorsed by ten of the interviewees, including three of the four managers, when asked about benefits of the appraisal system from the students' point of view. For example, one faculty member commented that, "if somebody was doing a crap job, they could be gotten rid of, or they could be taught, theoretically". Several others talked about the appraisal system as a kind of "quality control" that "ensured a high calibre of teaching and teacher". Similarly, one of the managers commented that "through the system of appraisal, it's kind of obvious if somebody has not been pulling their weight and that's how the students benefit".

6.2.3 Comparison with the literature

The findings cited above contrast quite sharply with the theory, so often expressed in the literature, that appraisal *ought* to improve learning, or, at the very least, teaching, more-or-less across the board, and certainly not just in the case of failing teachers. Thus, for example, Powney (1991a:172) contends that "improving the quality of the learning process is surely central in the objectives of the appraisal of teachers". Similarly, Mortimore and Mortimore (1991:127) argue that "appraisal ought to have an impact on the quality of student learning as well as the organisation skills, planning, and teamwork of the school staff", while Fidler and Cooper (1992:xiv) claim that appraisal "should lead to improvement in the learning

experiences of pupils and students". The same point is made in the 1991 DES Regulation, cited in Cullen (1997:181), which states that "appraising bodies should aim to improve the quality of education for pupils". A little less ambitiously, Magennis (1993:235) links appraisal to improved teaching rather than learning, writing that, "whatever the origin or stated purpose, the only reasonable justification for the existence of an appraisal system is the enhancement of the quality of teaching".

Nonetheless, my own research, in which 23 out of 29 teachers reported no change in their teaching, is by no means the first to point out a potential mismatch between the reality and the ideal of appraisal-driven classroom improvement. Turner and Clift (1988:173) write that "there was no shortage of evidence to suggest that little or nothing tangible seemed to have resulted [from the apprasial system]". Indeed, "it is clear that in many cases, appraisals did not seem to have much direct relevance for teaching techniques" (Turner and Clift, 1988:179). Other studies concur, suggesting that for a significant number of teachers, appraisal does not have any direct impact on teaching, let alone learning. Campbell (2002:160) cites two studies (one by Bennett, 1999, and the other by Wragg et at., 1996) in which 35% and 51% of teachers, respectively, reported that appraisal had no effect on their classroom practice; the figure for Campbell's own school was three out of eleven. Similarly, Kyriacou (1995:112) reports that "about two-fifths of the teachers felt the process had led to changes in their classroom practice", implying that 60% felt it had not. Cullen (1997:196) is more vague about the actual percentage, but her point, with reference to headteachers, is the same: "When asked about the effects of their own appraisal on the quality of teaching and learning in school and more generally on children, *many* of the heads thought that their own appraisal had little direct impact" (Cullen, 1997:196, italics added).

Thus, it would seem that, contrary to much of the rhetoric on appraisal, the causal link between monitoring teacher performance and enhancing student learning remains tenuous at best. Winstanley and Stuart-Smith (1996:68), cited in Campbell (2002:176), argue that, "There is no conclusive evidence

that the use of performance management systems results in improved performance". This certainly reflects the findings of my own study, where only about 20% of teachers reported any change in their classroom behaviour as a result of appraisal.

6.3 What connections, if any, do teachers make between student evaluation of teachers and grade inflation?

6.3.1 Findings from Rihab

Although teachers at Rihab were asked about the effect of appraisal on their classroom behaviour, they were not specifically asked about its effect on their grading of student work. Nonetheless, three teachers referred to this spontaneously. The first explained how, during double-marking, she and her colleagues would decide beforehand to give all the students A or B grades, even those who should have failed, because:

> You want to get good [SET] grades from your students because it has a major impact on whether ... you can stay in the country and work; if you can stay in the institution. So, in a way, you go out of your way to please the students. And you give the students what they want, and you give students inflated grades. You don't want students to fail and give you a bad evaluation.

Another teacher recounted how she had been reprimanded by management, ostensibly because she had been slated in her SET evaluations, after giving some of her students lower midterm grades than they had been expecting. Subsequently, she had inflated her marks, a policy she argued was covertly endorsed by the management, particularly in the case of students from well-connected families:

> And we were strongly given the message that the students' grades were going to have to be higher ... so that the students wouldn't be complaining ... so, I now started to get the feeling that if I wanted to survive at [Rihab] I would

have to give higher grades, especially to the girls [with royal surnames].

The third teacher had also received what she considered to be a negative evaluation and had then started "giving more opportunities for extra credit and feel-good things ... [allowing] a lot more leniency on tardiness and absences ... [and accepting] homework extremely late".

6.3.2 Findings from Al Fanar

As a result of what these three people had said at Rihab, I decided to modify the interview schedule at Al Fanar so that all faculty members were specifically asked if the appraisal system had had any effect on their grading of student work. All but one of them said quite categorically it had not. Interestingly, two of these people mentioned how it was not an issue for them, personally, but might be an issue for other people. One said, "in the States, they say that appraisal has begun to affect grading, and I think that it's an issue here, as well", whilst the other said, "non-Western staff [at Al Fanar] particularly have an awful lot to lose and might be tempted".

The one faculty member who did admit inflating 'discretionary classwork' or PDA grades said she did this because she had been told by her colleagues and even her manager that she would have some serious explaining to do if she gave a student low PDA marks (ones and twos) when all her other teachers had given her high PDA marks (threes and fours). Naturally, it could be the case that none of the teachers who allegedly told her to inflate her PDA grades were amongst my other interviewees. On the other hand, it could also be the case that they were, but chose to tell me one thing (that they did not alter their marks) and her another (that she should not give ones and twos). Obviously, I have no way of knowing which of these two scenarios is 'true', but this incident does serve as another illustration of the 'constructed' nature of interview data, and the way in which any conclusions drawn from them relate to reported perceptions of reality, rather than any notion of 'objective truth'.

6.3.3 Comparison with the literature

The literature on student rating of teachers is huge, particularly in the United States. It can be divided into two types. The first includes empirical studies that use statistical analysis of SET results and student grades in order to highlight correlations between the two. The second includes studies (both empirical and non-empirical) that use interviews, surveys and / or personal reflection in order to interpret these correlations.

The consensus from the first group is that students who achieve higher course grades give more favourable evaluations. Greenwald and Gillmore (1997:1210) point to "the widely observed phenomenon that course grades are positively correlated with course evaluation ratings".

The consensus from the second group is that this correlation exists because teacher effectiveness influences both student grades and teacher ratings. (See, for example, Marsh and Roche, 1997, and d'Apollonia and Abrami, 1997). In other words, good teachers are given better scores precisely because they help students learn more, and student ratings are a generally unbiased, valid and reliable guide to teacher performance "Dozens of scholars in the United States and abroad have agreed for years that student evaluations are a good measure of a teacher's skills. Nearly 2,000 studies have been completed on the topic, making it the most extensive area of research on higher education" (Wilson, 1998:2).

There are, of course, a few notable exceptions to this consensus, whose disconfirming claims are worthy of further investigation. Greenwald and Gillmore (1997), for example, offer three alternative explanations as to why higher student grades so often correlate with more favourable evaluations. Firstly, they suggest, student motivation (either in general, or in respect of a specific subject) could influence both grades and ratings. In other words, enthusiastic students might tend simultaneously to achieve high marks, and feel well-disposed towards their teachers, regardless of how well they actually teach. Secondly, students may infer that low marks are indicative of poor quality teaching, and, thirdly, they may 'reward' lenient grading

with high ratings. In these last two scenarios, grade-satisfaction is the critical intervening variable, rather than instructional quality or student motivation.

At the University of Washington, three studies on student evaluation were conducted using data from over 200 courses in each of several semesters, during the academic year 1993-1994. Based on statistical manipulation of the results, Greenwald and Gillmore (1997) concluded that the correlation between teacher evaluation scores and student course grades is best explained by grading leniency:

> Giving high grades, by itself, might not be sufficient to ensure high ratings. Nevertheless, if an instructor varied nothing between two course offerings other than grading policy, higher ratings would be expected in the more leniently graded courses.
>
> (Greenwald and Gillmore, 1997:1214)

Greenwald and Gillmore argue an elegant case, but theirs is still very much the minority voice, a fact that reflects my own research data, where 14 out of 29 faculty categorically denied inflating grades, and eleven made no reference to it. Obviously, the motivations of the four teachers who did admit to grade inflation are worthy of further study, but for our present purposes, it is enough to conclude that my study, like much of the literature, provides little evidence that student evaluation of teaching is seen as leading to grade inflation.

6.4 How much time do faculty and management devote to appraisal? Do they consider that the time devoted to appraisal is time well-spent?

6.4.1 Findings from Rihab

Unfortunately, interviewees at Rihab were not asked these two questions directly, so it is difficult to be sure how far they felt the demand for greater accountability was taking time away from other more important tasks, a

key faculty perception reported in much of the literature. Nonetheless, a number of faculty and managers did refer to the amount of time spent on various aspects of the appraisal system, and whether or not they felt the time was well-spent.

With regard to the teaching dossier, three faculty mentioned completing it "quickly", whilst another took "four hours" and a fifth took a weekend. A different faculty member mentioned how some of her colleagues "spent two or three weeks suffering" over the teaching dossier even though for her "putting the whole thing together maybe took me thirty minutes. It wasn't a lot of agony".

Several respondents (including both faculty and management) mentioned that the teaching dossier in the first year was more time-consuming to compile than the one in the second year, because the former asked for a statement of educational philosophy, while the latter required just a list of classes, achievements, and teaching materials. Some respondents said they thought the first year version was better because it "provided more valuable information"; whilst others said they preferred the second year version because it was simpler, more straightforward and quicker.

From a personal point of view, some members of faculty reported that the dossiers were useful, as tools for recording what they had done, reflecting on practice, and / or keeping up their CV. From a management point of view, however, four faculty members said they were "a waste of time" because the information they contained was unlikely to be read, let alone acted upon. Two of these faculty members mentioned how a colleague in a different department had deliberately submitted her first year dossier on a damaged / unreadable floppy disk and not been discovered, prompting one of them to ask, "Why should I waste my time writing something that no-one's going to read?". Whether or not management would endorse this sentiment is not clear, but the senior manager to whom the dossiers were turned in did admit that she was unable to review them in any great detail because there were 105 of them to be read every year.

In summary, it seems that faculty did not spend a great deal of time on the appraisal system, but, even so, there was a certain amount of resentment over the small amount of work that was required, because management were seen as ignoring the data they demanded. This sentiment is best summed up by the following faculty comment:

> I think that preparing dossiers and other kinds of hoops that we've been asked to jump through by the administration, in the name of evaluation, and assessment and whatever, has impinged on our time, but it's usually anywhere from an hour to a week of preparation. You know, it's not too much to ask, if they're used properly, but I haven't really seen that they've been used. Like I said, when the rehire decisions are made, then I guess we'll have a better idea of the thing being used. But I don't have much faith that they're going to be used.

With regard to the managers themselves, two of them admitted that whilst the appraisal system took up a great deal of time, it was not particularly effective, and yet, paradoxically, making it more effective would take up even more time. The first said:

> For something which is not very standardised, not standardised enough to draw any general conclusions across the university or even across the unit, it takes up an inordinate amount of time. On the other hand, to make it more worthwhile, it would have to take up even more time. If we're going to do it properly with interviews and consultations and so on, it does take up a long time.

And the second said:

> The drawbacks are just the paperwork. With 105 faculty, think of the hours ... I guess I just don't see it as terribly meaningful. If I had ten faculty members, fine, that'd be great. I could spend a lot of time and go and talk and things like this. But there are so many people in this case, you

know ... so, it's fairly mechanistic, if you will. I suppose ... [it could function more effectively] ... if the instruments were better, and more time were spent with the individual teachers in talking and discussing ... [but] any hours not in class, you really have to justify, and so, more hours spent on these kinds of things means more hours out of class.

Both respondents suggested that this allocation of additional time was unlikely to happen because appraisal was not a priority, in the institution's second year, when so much else was still at the developmental stage:

> I think, at the moment, it's a token effort. The authorities realise that they need to have some system of staff appraisal and most organisations these days do have some system of appraisal intended in a general sense to make the organisation as efficient as possible. But to do that in the ideal world is going to take far more time than we've got. And the present system is a stopgap, and I think it's recognised that it's inadequate but as I said before, it's not a high priority when we've got a lot of work to do on curriculum and standards and all the rest of it.

This lack of priority was also referred to by another manager, rather poetically, over lunch one day. According to my field notes, in the context of a discussion over appraisal, she recalled how, in the university's first year, "water" was used "to put out the fires in the administration block", rather than "to tend to the flowers in the garden".

Management therefore seemed to be aware that the appraisal system might be taking up more time than its impact justified, but were allowing this to continue, at least in the short-term, because other issues were more pressing. It remains to be seen whether this situation changes as more of the university's inaugural work on curriculum and assessment becomes routinised.

6.4.2 Findings from Al Fanar

After analysing the data from Rihab, and identifiying time as an important aspect of appraisal, I modified the interivew schedule used at Al Fanar so that interviewees were specifically asked, firstly, how much time they had devoted to the apprasial system, and, secondly, whether they considered this time well-spent. The first of these questions threw up an interesting range of answers because some faculty defined the appraisal system quite narrowly in terms of the class observation, the student evaluation of teachers, and the completion of a self-evaluation form, whereas others interpreted the appraisal system more widely, and included in it time spent in professional development workshops, time spent "thinking about my work while I'm watering my flowers at home", and even "anything that you do apart from the bare necessities of our job, which is preparing lessons and delivering lessons".

The time spent filling in the self-evaluation forms, preparing for the class visit, and reading the results of the student evaluation of teachers was generally defined in terms of a few hours, and considered by almost all faculty to be "very little" or "not much".

With regard to management, all three middle managers mentioned how time-consuming it was. One had even given up two weeks of her holiday in order to complete the summative evaluation forms on the teachers she managed. The senior manager also commented on the fact that although the appraisal process was not particularly time-consuming for her, it was "bloody time-consuming" for her middle managers. Many of the faculty members also recognised this. One probationer mentioned how she had only been observed once, not twice, because her manager was so busy; and another commented on how "people feel it [appraisal] is the right thing to do, but it's always given to the people who are busiest, supervisors, middle managers ...".

In response to the second question, nine of the fifteen faculty members reported that the time they spent on appraisal, whichever way they

calculated it, was time well-spent, either because they "enjoyed" doing it, they felt it provided their manager with useful information, or it helped them to "take stock" and reflect on their own practice. One of these nine did mention, however, that although she regarded her own PD activities as worthwhile, she thought some of her colleagues were devoting energy to "projects that aren't necessary and are a waste of time and effort", just because they felt under pressure to set and achieve some goals each year. She noted the importance of linking one's own professional goals with the needs of the institution, adding that she had worked at another college in the system and felt that people there, in contrast to Al Fanar:

> certainly produced reams and reams of booklets and exercises that no-one ever cracked open again. It was just wasted paper really. But they said they were going to do it, and they just had to produce something at the end of the year.

This need to marry the goals of the individual with the needs of the institution was also mentioned by two of the managers, one of whom sent out a list of departmental goals at the same time as people were asked to complete the goal-setting section of their self-evaluation form.

By contrast, four of the faculty members suggested that the time spent on appraisal was "a waste" because they did not learn anything as a result; it was perceived as just another "admin task we have to do". The remaining two faculty members answered the question in such a way that it was not clear whether they thought the time was well-spent or not.

Obviously, it is hard to compare Rihab and Al Fanar faculty reactions on this point because the former group were not specifically asked about it, but my sense is that faculty at Al Fanar were generally happier with the time they devoted to the appraisal system, because the majority of them said the time was well-spent, and the four people who did say it was a waste of time did not express these sentiments as strongly, nor at such great length, as the four people at Rihab.

With regard to the management at Al Fanar, all four interviewees stressed how important appraisal was and there was no suggestion that they resented the amount of time that had to be devoted to it. The senior manager, having noted it was "bloody time-consuming" for middle managers, added "but that's fine ... it's a very important part of their job, if not the most important part". Similarly, the middle managers themselves called it "hugely important", "incredibly important" and "one of the things we should be spending time on". Two of the managers were even planning to devote yet more time to it in the coming year, one because she wanted to make it "less clinical", and one because she wanted to make it "more developmental".

One of the managers did comment on how appraisal always gets relegated, "to the bottom of the pile" behind student problems, grade issues, and exams, but, in general, the managers at Al Fanar reported a rather stronger commitment to the appraisal process than those at Rihab, perhaps reflecting the fact that Rihab was only two years old and still very much in the process of developing curricula and assessments, whereas Al Fanar was five years old, and part of a well-established system more than ten years old.

6.4.3 Comparison with the literature

The findings cited above, particularly from the faculty members, are somewhat at odds with much of the literature, where considerable concern is expressed over the amount of time that appraisal requires. Many teachers in the study by Turner and Clift (1988) complained that not enough time was allocated to appraisal, with the result that other activities had to be curtailed. Likewise, amongst teachers in Kyriacou's (1995:114) study, "the most common complaint" was that appraisal was "very time-consuming", whilst in Campbell's (2002:161) study, "lack of time was regarded as the most difficult obstacle to overcome".

Similarly, amongst the headteachers interviewed by Cullen (1997), "the overwhelming majority" (Cullen, 1997:191) "found the appraisal process

very time-consuming, and as a result experienced some stress as well as concern that time was being taken away from other important school activities" (Cullen, 1997:200).

In Higher Education, the same concern is even more apparent, though here it is most often linked to accountability in general, rather than appraisal systems, per se. Trow (1994), Randle and Brady (1997b), Elliott and Crossley (1997), and Currie and Vidovich (2000) all mention the increase in bureaucratic procedures, burdensome paperwork, and form-filling that has come about as a result of managerialism, more specific details of which have already been discussed in the section on the relationship between managerialism and bureau-professionalism.

In comparison to the studies cited above, the data from Rihab and Al Fanar appear rather more complex. The system of appraisal at Al Fanar was considerably more elaborate than at Rihab, involving, as it did, individual faculty interviews and summative reports. Nonetheless, managers at both institutions mentioned how very time-consuming appraisal was, and, in this respect, my own findings concur with much of the previous research. With regard to faculty perceptions, however, the position is somewhat different. In the literature, both managers and faculty seem equally concerned about the amount of time appraisal requires, whereas, in my own study, this was mentioned far more often by managers than by faculty. Indeed, very few faculty made any reference to the time-consuming nature of the appraisal process, and those that did were concerned, not so much because they had more important things to do, a common complaint in other HE studies, but because they believed very little use was being made of the information asked for.

6.5 To what extent is appraisal said to focus on either basic technical competencies or flexible, creative practices?

6.5.1 Findings from Rihab

Although none of the interviewees at Rihab was asked directly about this dichotomy, all the managers were asked why the specific appraisal methods used at the institution had been chosen, and all the faculty were asked whether they agreed with the methods adopted. Opinion was very much divided over whether any of the three appraisal instruments (the teaching dossier, the lesson observation, and the student evaluation of teaching) were of any value, but there was certainly no consensus that the instruments focused on mechanistic number-crunching or technicist performance indicators.

Two of the five managers said the teaching dossier was the most "important" or "valuable" element in the appraisal process, although a third called it "a formalised waste of time" because it was written during two weeks in April instead of being an on-going process throughout a teacher's period of employment. One of the managers said she preferred the dossier from the first year because it required a detailed statement about teaching philosophy, rather than a "perfunctory" list of classes taught, whereas another said she preferred the scaled-back version because it was less time-consuming for faculty to write and appraisers to read.

Faculty, likewise, were divided over whether the dossier was worthwhile, and whether the first or second year version was better. One faculty member said she preferred the earlier version because it was "more personal and more teacher-specific", whilst another said the "simplified" version was a less "ambiguous" improvement. Four faculty suggested it was a waste of time because it was not going to be read and would have no impact on their future employment. Two of these faculty recounted the anecdote, mentioned earlier, about someone in another department who deliberately submitted her dossier on a damaged / unreadable disk and was never asked for another copy. Six other faculty noted how, in the second

year, at least, the dossier had been read by someone, since they had received a letter from their Head of Department that made some reference to the material contained within it. Interestingly, two faculty maintained that the dossier would only be read in detail if a manager were looking for evidence to use against a faculty member. Regardless of whether or not management read the dossiers, five faculty said it was personally useful for them to have compiled theirs, either because it helped them update their CV, or because it made them feel good about all they had achieved during the year.

The general consensus therefore seems to be that the teaching dossier was a potentially useful document, particularly if it were read by management and used to inform their decision-making. As was suggested earlier, with one possible exception, faculty did not seem to consider the dossier either mechanistic or technicist.

Similarly, although most faculty said they gained little, if any, benefit from the classroom visit, there was no suggestion that the categories on the lesson observation form were inappropriate.

Two faculty members described the classroom visit as "perfunctory", and another recounted how her observer had spent half the time checking her email. A fourth faculty member recalled how her observed lesson had been entirely taken up with student presentations. The same thing had happened during one of my own lesson observations, and in my case, the observer had justified this kind of class visit by saying that lesson observation "was just a formality".

Despite the fact that pre- and post-lesson discussions are usually considered an integral part of lesson observation, at least according to the extensive literature on appraisal, no-one mentioned having a pre-lesson discussion, and only one manager and three faculty members mentioned having a post-lesson discussion. One of the managers said she did not have time for post-lesson discussions and felt some teachers would resent them. Interestingly,

this sentiment was not reflected in what the faculty themselves actually said. Understandably, none of the teachers said they would resent a post-lesson discussion, but, more surprisingly, four of them said they actually wanted such a discussion, and were unhappy about not getting it.

With regard to feedback, one of the managers openly admitted that she only ever gave top marks and wrote positive comments on the feedback form, because anything else "could so easily be misinterpreted" at a later stage. She might make suggestions for improvements orally, but these would not be included in the official form because "developmental practices, I feel, are better done, off paper". Two of the other managers said that they did use a mixture of ones ("needs improvement") and twos ("does well"), though, according to faculty responses, very few ones were ever recorded. Indeed, all but one of the faculty who talked about their numerical feedback said they got top marks in every category. For one person, this was a fair reflection of an experienced, well-qualified staff. In her words, "we're qualified, experienced teachers, we should get twos across the board". For many others, however, this strategy was unhelpful. Five faculty complained about the fact that they had not received any kind of detailed feedback, saying they would have positively welcomed suggestions on how to do things differently. Their sentiments are best summed up by the faculty member who commented:

> Basically you get a score of one, two or zero, and everyone just gets two, two, two, two, two. And I've never seen anybody that hasn't. And that's completely useless, as far as I'm concerned, because, when I do a lesson, and I'm observed, I don't want to be told that that was a perfect lesson, and I got two, two, two, two, two. I want to be told, okay, what the observer thought went well, but areas where they thought I might want to work on. There are always trends that you can work on.

One member of faculty did recognise how observers might be reluctant to note down areas for improvement on the official form, because it became

part of your permanent record. In this case, she suggested they give faculty members a second sheet of handwritten notes detailing what they really saw and thought could be improved.

There was just one instance where a faculty member said her lesson observation had been a valuable experience. In her case, the observer had conducted an extensive post-lesson discussion and had given her a feedback form that contained a list of strengths and weaknesses, rather than specific numbers.

Thus, it seems fair to conclude that whilst faculty generally found the numbers on the classroom visit form unhelpful, they were not unhappy with the focus of the observation, just the lack of constructive feedback.

The picture was a little less clear with regard to the student evaluation of teaching (SET). Four of the five managers downplayed the importance of the SET questionnaire (and the fifth stressed this in a memo sent to all faculty, though not in her interview). One said she "disregarded them", and another said:

> I don't take it terribly seriously given the way it is, the way it's conducted and everything else. I should, I suppose, and I suppose people above me would be appalled if I said – but I have said this in meetings, that student evaluations I don't take seriously at all. No-one here, at least in one year's time, would ever be dismissed, as far as I'm concerned, because they had poor student ratings.

When asked why she had come to this conclusion, she replied:

> These students aren't used to doing things like this at all. And I look at the comments they make and so forth and they're very echoic. They say the same things and I think they're essentially meaningless.

Later, she added that student evaluation results would only catch her attention if a person consistently got the lowest ratings of any teacher over a period of three or four years. This need for cumulative evidence was also stressed by two other managers. One suggested questions should be asked if a person received negative student evaluations "over three consecutive semesters", while the other expressed the same sentiments, though she was less specific about the time-frame:

> I think the student evaluation over time would probably give a good impression, but for teachers who've only got two classes in each semester, maybe four classes in the year, it's going to take a while for a realistic picture to appear, because it's inevitable that you'll get classes that you don't get on with particularly well. So I think it's more important for the students that they feel they have a chance to comment on the teachers and I think on the whole they do it quite responsibly.

Faculty opinion was more divided on the issue. Six faculty members mentioned how important it was to allow the students to give feedback, and one of them went as far as to suggest that their feedback was the most valuable, firstly because they saw the teacher every day whereas the manager visited just once or twice a year, and secondly because:

> They're our clients. So that's who we should be doing things for. If they're not happy, then we're doing something wrong.

At the other end of the spectrum lay several faculty who argued that the SET questionnaire was not a valid instrument, either because it was badly designed or because the student population was not sophisticated enough to use it properly. One teacher complained that:

> because the student questionnaire evaluates the course at the same time as the teacher, you get negative evaluations based on things that are completely out of your control.

Two other teachers suggested that some of the questions, particularly one about setting homework every day, were not relevant to what a good teacher does.

A further two teachers said some questions on the SET questionnaire were far more important than others, even though they were all given the same weighting. One of them then went on to express concern that the figures for all the questions were averaged to yield an overall numerical rating for each teacher:

> I think that it [the SET questionnaire] is appallingly ill-designed, and very often provides wrong information, I mean, at the most basic level. I give you one example – questions are averaged, so you get a one, two, three, four, five. And answers are assumed to be good or bad. Like your teacher gives you homework every day is assumed to a good thing. But certainly in the programme that I've been working on 50% of the teachers would never give homework, because their job has been specifically to supervise the homework that is given by another teacher. So you can't possibly get any meaningful information from that. And when that is averaged, and you get a 2.1 or a 1.2 or whatever the number is they're looking for at the end, it just skews things and the whole thing just becomes meaningless.

In addition, four other teachers argued that the students did not complete the questionnaire appropriately because they were too "finicky", "whimsical" or overly influenced by how well they were doing on the course.

Several faculty complained that the SET results were just handed to staff without any discussion. They argued that the numbers in themselves "don't seem to mean very much" and classroom behaviour "is not going to [change] if you don't get any kind of feedback other than a few numbers". Two faculty members had taken it upon themselves to address this concern

by soliciting their own on-going qualitative feedback from students. According to one of them:

> I'm not really very interested in the form-filling evaluation. I'm much more interested in the student evaluation of our work together, and I get lots and lots of feedback on a regular basis. I mean, I do talk to my students, daily, weekly, about how **WE** are doing, about how **WE** are learning together. I talk to them about how they want to learn, what they want to learn, why they want to learn.

Overall, it seems that the faculty were quite dissatisfied with the current SET instrument, for a wide variety of reasons, some of which relate to a perceived over-emphasis on numerical values and behaviour that is easily observable but not necessarily indicative of good teaching. Interestingly, the SET questionnaire was redesigned in its third year and my penultimate interviewee mentioned how she thought the latest version was better because it focused systematically on enthusiasm, rapport and organisation, which, in her eyes, were much more relevant to effective teaching than the giving of homework.

6.5.2 Findings from Al Fanar

At Al Fanar, faculty were asked whether they agreed with the methods of appraisal that had been chosen, and whether they thought the instruments focused on appropriate criteria. In general, the faculty at Al Fanar spoke of a need to judge people according to explicit criteria, in order to be as objective and therefore as fair as possible. In the words of one middle manager, "it can't be subjective, vague or wobbly". In the words of another, it needs to be "objective ...[and] ... evidential". Somewhat surprisingly, the most senior manager asserted that the process should rely 90% on what is observed and 10% on "gut reaction", suggesting that if it were 100% objective, it might not be so good.

With regard to the specific criteria for appraising faculty, seven of the teachers at Al Fanar said they were happy with the methods used, which, in

their words, were "fairly standard", "fairly professional", "fairly carefully thought-out", "quite good", "fair" and "okay". Four of the remaining faculty mentioned the importance of considering the methods of appraisal within the wider context of the organisation. In other words, the appraisal methods used at Al Fanar had the potential to yield useful information, but this potential would be realised only if the general atmosphere was one of openness and trust. In the words of one interviewee:

> I think that some of the things may or may not be valuable, but if you haven't got confidence that there are procedures that are to be followed with everybody equally, and decisions are open to public inspection etc. etc. – if none of these factors are there, it doesn't really matter what procedures you use.

A further two faculty mentioned how the appraisal system created "a veneer" of following "best practice", even though nothing much changed.

Almost all faculty said that when they were being observed, they tried to give a lesson showing themselves in the best possible light. "People do get the impression, rightly or wrongly, that it's a formulaic lesson they are looking for." Indeed, one new faculty member recounted how, after giving "a performance lesson" for her first observation, she had given "a regular lesson" for her second observation, and been told this was not quite right. The manager observing her had asked for a more detailed lesson plan, and other faculty had told her, "You're not supposed to do regular lessons. You're supposed to do performances ... when you are being observed you do a special observation lesson. It's not what you normally teach".

This emphasis on performance was also apparent in the comments from managers, one of whom said she told her faculty "it'll go in your evaluations, so yes, do an all-singing, all-dancing [lesson]". Similarly, the senior manager likened a class observation to hosting a dinner party where "I'm not going to bring out my broken crockery and cups – I'm going to get the best stuff" meaning that observed teachers should give their best

possible performance. Clearly, neither teachers nor observers saw the class visit as an opportunity to engage in action research.

There was much less of a consensus over the appropriateness of the criteria used to evaluate the observation. One person maintained that the form was good because it focused on "how you are as a teacher rather than how you were in that particular lesson". Four other faculty disagreed, saying the form needed to be "more general" or "more flexible" so it could be used with a wider variety of lessons, and branding some of the categories "dingbats, daft, simplistic" and "ludicrous". A further two faculty members mentioned how the form seemed to place too much emphasis on what was currently "fashionable" in English language teaching. One faculty member suggested that the criteria were "insultingly simplistic" and should be done away with, although then "they [the observers] could stand accused of being too subjective".

One faculty member distinguished between "surface" criteria such as "classroom control and neatness" and "significant" criteria such as "is the teacher attending to all the different levels of students involved in the classroom". She said the current form contained a mixture of surface and significant criteria. This distinction between what was easily observable and what was really significant was also mentioned by two of the managers. One argued that the criteria "didn't really focus on important classroom skills. They tended to focus on peripheral things, such as writing up the objectives on the whiteboard". The other made a similar point, saying:

> It's got these horrible boxes ... the form itself is actually very judgemental ... it's such an obvious ticklist. You know, so how many ticks did you get, or how many fives did you get equals a good lesson, you know. So, I mean you could have a robot in there, couldn't you? You know, wrote up the objective on the board, you know, stated the objective clearly, got them to do the work and then left. Like, where was the interaction, where was the communication? Where

was anything in that? So, yeah, it's very basic, it brings it all down to a very basic level. I find it really nasty.

On the other hand, both of these managers were aware of the danger of being seen as "subjective" in their judgements.

Accordingly, both faculty and managers seem to perceive the classroom visit as focusing, at least in part, upon those features of the lesson that are most easily observable, but only in order to reduce the apparent subjectivity of the process. In other words, it was not accountability, per se, that led to an emphasis on basic technical competencies; it was, instead, a desire to ensure as much objectivity as possible.

Some faculty appeared happy to receive a numerical rating for their lesson observation because "being given a ranking happens everywhere" and "you expect to have your skills ranked". Others felt rather more uncomfortable with this because "it's hard to make people into numbers". One of the managers, in particular, was adamant that "actually having to give a number to what somebody is doing in the classroom" was "quite difficult" and "not particularly constructive".

Similarly, one of the teachers recommended that student pass rates be used to determine teacher effectiveness, whilst another argued the exact opposite. The first teacher said student pass rates were more "objective" and less "political" than other methods, such as the manager's summative report. By contrast, the second teacher said "it's really hard to measure efficacy in teaching and learning numerically" because students might make "a great deal of progress, and a lot of growth, and a lot of personal development" even though this was "not reflected in the pass rate".

With regard to the student evaluation of the learning environment (SELE), I found the same spectrum of opinion as at Rihab. One teacher suggested that student feedback was the most valuable kind because they saw you "180 days a year"; many other faculty members dismissed the results as

"rubbish", "a mess", "bizarre", "nonsensical", "a beauty contest", "a farce", and "worthless". Amongst this group were three teachers who each recalled having a public disagreement with a particular student in class and then being slated in their SELE shortly afterwards. Between these two extremes lay a few faculty who said student feedback could be useful if the questions were suitable, and the students were given sufficient time and training to allow them to respond constructively. Two of the managers concurred with this viewpoint, and suggested that comments from students were more informative than mere numbers.

One of the managers argued that the latest SELE version was "less subjective, and ... less [of] a tool to attack a teacher" whereas another maintained that it was still fairly unhelpful because it focused on peripheral things, like teacher punctuality, when "there are more important things we want to know".

So again there is some suggestion that the SELE does indeed focus on the most easily observable facets of teacher behaviour, but again, this is seen as a legitimate attempt to limit the subjectivity of the instrument, an apparently prudent precaution, given how many teachers complained of students acting on a "whim" or holding "a grudge".

The final element in the appraisal process at Al Fanar was the annual setting of individual goals, something that had no parallel at Rihab. Three faculty said they welcomed this as an opportunity to take stock and inject some direction into their careers. On the other hand, a couple of faculty said self-evaluation could not be completely honest at Al Fanar because people wanted to protect themselves, rather than admit to a genuine weakness. One of them remarked how:

> when you have an organisation that you feel completely safe in, then I think those self-evaluations can be particularly enlightening. But, what happened with us was we got all these rumours ... that a bunch of people were going to be laid off, not make probation, so that takes you away from

> being particularly honest into being sort of protective. I suggest that everybody wrote stuff, not because they were doing self-evaluation, but they were trying to answer the questions correctly so that they would be able to continue here.

A third faculty member also suggested that the self-evaluation process encouraged people to engage in activities just for the sake of it. In her opinion, some of her colleagues, particularly at another college, were wasting their time on unnecessary projects, just so they had something to write on their self-evaluation form. In this case, the need for greater accountability may well have led to some spurious developmental activity, though, of course, such an opinion is hardly representative of what the other fourteen faculty members thought.

6.5.3 Comparison with the literature

A number of authors have commented on how most appraisal schemes are competency-based and as such limit teaching to a narrow set of observable skills. According to Peaker (1986), appraisal procedures in the US encourage "safe" rather than creative, flexible teaching, focusing on highly specific competencies at the expense of a more holistic assessment of a teacher's overall contribution to the institution. Similarly, Wragg (1984) and Elliott (1983), cited in Turner and Clift (1988), suggest that competency-based appraisal in the UK ignores the need for imaginative and reflexive skills. More recently, Troman (1996:22) has argued that contemporary definitions of a 'good' teacher place greater emphasis "upon the technical competencies that facilitate the administrative and management aspects of the ... teacher's work" with the result that "teacher quality is now defined in terms of technical competencies as opposed to personal qualities". Quality is reduced to observable facets of student learning and only the predictable and measurable is valued, making what is non-observable, non-quantifiable or unintended irrelevant (Elliott and Crossley, 1997:84). The more teacher appraisal is concerned with "managerial control", the more quality will be "defined in terms of minimal

competencies ... [meaning] ... the most generic of technical skills" (Ko, 2001:35).

Without doubt, aspects of the appraisal schemes at Rihab and Al Fanar were said to focus on basic technical competencies. Particularly at the latter institution, the observed lesson was described as an opportunity to perform at one's best rather than as a chance to engage in collaborative experimentation. There was also some concern on the part of both faculty and managers about the reduction of the complex teaching process to a series of numbers. However, for the most part, teachers suggested this was preferable to anything more nebulous and open to subjective interpretation, and there was certainly little evidence that teachers linked either the use of numbers or the focus on basic technical competencies to a desire for greater managerial control.

With regard to student evaluation of teachers, the degree of diversity found at Rihab and Al Fanar is also present in the literature. Some authors, such as Pollitt (1987:96), place considerable value on the judgement of learners they label "consumers". In their eyes, consumer accountability can complement professional development, and empowering students in this way can only enhance the work of teachers. Similarly, Mortimore and Mortimore (1991:131) suggest that "clients" have valid opinions and perspectives, although they do acknowledge that they may not always be in a position to make informed judgements about a teacher's work. By contrast, authors such as Trow (1994:15) and Trowler (1998a:105) argue strongly against demand-led teaching, which they believe short-changes rather than empowers students, because it is a teacher's job not merely to fufil students' expectations, but to challenge, modify and extend them.

Many of the teachers at Rihab and Al Fanar agreed with the principle of allowing students to have their say, but disagreed with the particular instrument that was used. A similar dichotomy between theory and practice has also been noted by Boyd et al. (1994:12), who write that:

A large majority of professors in most countries agree that student opinion should be used in the evaluation of teaching effectiveness ... Still, we find widespread dissatisfaction with faculty evaluation. Replies from all countries overwhelmingly indicate that better ways are needed to evaluate teaching performance.

Such dissatisfaction with student evaluation of faculty has also been previously noted at tertiary level in the country in question. According to the results of a survey conducted by Richards (1999), 23% of faculty at a sister institution to Al Fanar were dissatisfied with SELE as a means of evaluating their performance, and 38% were very dissatisfied. Many teachers in this study said students at the college were too immature to make an objective judgement about their teachers; they expressed concern about students holding grudges; they mentioned how students confused popular and good teachers; and they suggested stricter teachers got lower ratings; all of which echoes the findings of my own research.

Moreover, with regard to target-setting, Campbell (2002:75) reports that many of the teachers in his study found this to be the most difficult aspect of the appraisal process to get right. This does not seem to have been the case in my own study, although a couple of people did mention a certain reluctance to admit to weaknesses, and a third person suggested target-setting might lead people to engage in 'busy work' for its own sake.

Finally, the tension between objective and subjective criteria, mentioned by several managers and faculty, has been widely discussed in the literature. On the one hand, objective criteria, applied to everyone equally, may be the only way to ensure the evaluative element of appraisal is fair, rather than biased; on the other hand, subjective criteria, tailored to individual circumstances, may be the only way to ensure developmental appraisal is meaningful, rather than trivial.

Accordingly, the primary school heads in Hellawell's (1991:34) study reiterated the need for both standardised procedures to ensure equity, and differentiated procedures to ensure individual development:

> In essence, the majority of heads appeared to be saying that they wanted universality of procedure to some extent to ensure parity, but they also wanted particularistic treatment to ensure their own peculiar circumstances were taken into account.
>
> (Hellawell, 1991:35)

Similarly, Gleave (1997), cited in Campbell (2002:21), suggests precise, objective assessment is necessary when performance appraisal focuses on accountability, whereas flexible, subjective interpretation should be utilised when performance appraisal focuses on development. Would that it were that simple. In reality, many schemes focus on both evaluation and development, and many seemingly objective criteria are actually a matter of interpretation, because "judgements of performance are necessarily value-laden" (Simons and Elliott, 1989:189).

Interestingly, both management and staff at City University of Hong Kong spoke of the need not to limit academic appraisal to quantifiable outputs (Ko, 2001). Amongst both groups, there was said to be support for subjective judgements, provided these were not used to make too fine a distinction between the performances of different faculty members. The senior management, in particular, "seemed to have a shared view that it was not only *impracticable,* but also *undesirable"* to restrict measurement to what is "quantitative, precise, or predictable" since this "would mean that performance would be mediocre, and not creative enough" (Ko, 2001:156, italics in original). In other words, "one had to accept subjective judgements in assessing academic work ... [because there was] ... no better way to assess academic work" (Ko, 2001:156).

Very little support for this kind of subjective judgement was found in the current study (with the possible exception of the senior manager and a few

faculty members at Al Fanar). On the contrary, the need for objective criteria and clear evidence was repeatedly emphasised, and the lack thereof repeatedly criticised – a situation that may well be related to the lack of job security and prevalence of politicking mentioned by many interviewees, and discussed in more detail in sections 6.7 and 6.8 below. It seems reasonable to assume that at institutions where fairness and parity are major bones of contention, staff are likely to prefer the protection seemingly afforded by transparently objective criteria and extensively documented evidence; at institutions where staff already feel secure and equitably treated, they are likely to accept, and even embrace, the developmental possibilities seemingly facilitated by more subjective, individually-tailored criteria.

6.6 How far is the specific appraisal system at each institution perceived as developmental and / or evaluative? How far is it seen as embodying a paradigm of professionalism and / or managerialism?

6.6.1 Findings from Rihab

In relation to this question, interviewees at Rihab were asked firstly, if they thought the appraisal system at the institution was primarily evaluative or developmental, and, secondly, whether they thought it was possible and / or desirable to separate these two aspects of appraisal. Analysing the results of this section was particularly difficult, partly because some respondents appeared to answer these questions in ways that contradicted assertions made elsewhere in the interview, and partly because merely equating development with professionalism, and contrasting this with evaluation / managerialism fails to do justice to the intricate relationships between these four elements.

Certainly, the great majority of interviewees reported that the present system was more evaluative than developmental. This was the view expressed by three managers and eleven faculty members, although five of this latter group then went on to say that, in reality, it was neither

evaluative nor developmental because nothing was done with the information collected. In the words of one interviewee, "There's no purpose to this. It's just an exercise, because there's no feedback. It's just something they have to go through". Another interviewee commented, even more pointedly, that:

> The two main functions, in my opinion, of faculty appraisal – number one would be for the administration to figure out who to retain, and number two, would be for the faculty to improve themselves. And on both counts, I don't see our appraisal system really having much bearing ... It doesn't seem to have much impact. There don't seem to be any decisions being made based on the information ... .

In similar vein, various other interviewees labelled all or part of the appraisal system as "perfunctory", "fairly mechanical", "window-dressing", "a waste of time", "quite, quite shallow", "ad hoc" and "very, very limited".

Interestingly, although many of the interviewees were unhappy with the current system, they did not express any objection to evaluation, per se, and nor did they point out any inherent contradiction with combining both evaluative and developmental elements within the same system, contrary to much of the literature on appraisal. To be fair, one manager did say that appraisal could be either evaluative or developmental, but not both simultaneously. By contrast, two managers and seven faculty members said that even though the current appraisal system was primarily evaluative, it would be better if it also included more developmental opportunities. Several faculty members reported that they would prefer more detailed feedback, particularly after the classroom observation, in order to develop professionally.

One talked about "missed oppportunities for development" and another said:

> I feel like I have a lot to learn from people who have more

> experience than me. And if I have to go through the pain of observation, I'd a lot rather actually get something out of it. Rather than a piece of paper that says everything is fine and one verbal comment that says well, it wasn't the most exciting lesson I've ever seen, but it was okay.

Somewhat surprisingly, just over half the faculty members (but none of the managers) said that as well as failing to provide opportunities for professional development, the appraisal system was also very negatively skewed. Four of them claimed that it offered no positive recognition of achievement, and a further four suggested that it was used only as a tool for getting rid of people. Their comments included the following:

> Some of the teachers here in particular, I think, are doing a fantastic job; they don't get any kind of rewards for that.

> There's just no recognition whatsoever, certainly no positive recognition of anything anyone does. Presentations at conferences, papers published, other kinds of professional advancement, there's no recognition at all.

> You [the management] say – we're just trying to check that you're not doing something bad. We're not looking at, you know, what are you doing that's good.

> I have never heard of, or seen anything, that showed any sort of positive reinforcement if one does well .... it's just a negative.

> And the feeling – I think it's a pretty general feeling – is that any appraisal of you that happens only matters if they want to get rid of you. If they want to keep you, nobody is really going to pay any attention – it's not going to help you.

In addition, six faculty members spoke disparagingly of a management that are "downright incompetent"; that are "horrible and mean and spiteful and lacking in trust"; that lack knowledge of basic management principles; that send emails "full of shit" and make "asinine pronouncements [every] two or three days"; that "stomp on everybody to keep them down"; and use teachers as "cannon folder to shore up the various holes that appear every time they have a new idea".

One, when asked what effect the appraisal system had had on her relationship with the university management, replied that it had, "continued to confirm my opinion that they do not know what they are doing". Rather more strongly, another described the faculty appraisal system as "just another policy ... more outrageous bullshit coming down the pipe [from the administration building]".

Obviously, given my own situation, I feel very negatively about the appraisal system at Rihab. Throughout the data analysis stage of my research, I was very conscious of the need not to let my personal experience prejudice the results, and not to focus unduly on comments that reflected my own feelings. For this reason, I systematically re-analysed the data in this section, specifically looking for any positive notes. I found none, and therefore feel justified in claiming that the negative comments cited above are a fair representation of the views expressed by around half the faculty. In section 6.1, it was argued that faculty are not opposed, in principle, to accountability models of appraisal. Accordingly, it would seem that those who spoke negatively about appraisal at Rihab did so, not so much because they saw it as primarily evaluative, but because they described it as being poorly administered by a management they did not respect. Such a stance was explicitly adopted by two interviewees who commented that:

> I do think it's a necessary evil. I just think it's done badly here.

And:

> I think it's necessary – for any institution to function, it has to have some system of evaluation. But that doesn't mean to say that I approve of this one.

6.6.2 Findings from Al Fanar

In order to try and elicit less confusing data, the interview schedule at Al Fanar was modified slightly. Instead of getting interviewees to make a bi-polar choice between evaluation and development, I asked them to plot the college's position on an evaluative / developmental continuum, thereby allowing them a better opportunity to expand on their perspective and make more finely-graded distinctions. I also asked them how the Al Fanar system compared with any appraisal system they might have experienced elsewhere, in an attempt to separate opinions about appraisal in general, from perspectives on the specific system in place at Al Fanar.

Overwhelmingly, faculty placed the system towards the evaluative end of the spectrum. Only one said it had been developmental, in that she had written an area of interest on her self-evaluation form and subsequently been allowed to concentrate on that area. Two other faculty members said that although it was currently evaluative, in the past, with different managers, it had been much more developmental, as they had been given a more detailed critique of the observed lesson containing suggestions for improvement. They both thought the precise balance between evaluation and development was more a function of a particular manager's approach than a function of the system itself, echoing Cooper and Fidler's (1992:359) contention that "An appraisal system is only as good as the managers who have to operate the system".

The rest of the faculty members placed it firmly towards the evaluative end of the spectrum, eight of them suggesting it was mostly evaluative, and a further three suggesting it was "wholly" or "95%" evaluative. Two of the managers placed it firmly "in the middle" whilst the third said it was "fairly high towards the evaluative and 'judgemental" end of the spectrum, especially when compared to the very developmentally-focused system she

had used the previous year in her former institution. She went on to say, somewhat hesitantly, that she suspected senior management would not like publicly to admit how evaluative the system was because they were aware that this was not considered "best practice" in education. Interestingly, within the context of the confidential interview, the senior manager had no hesitation in declaring the apprasial system "90% towards the evaluation and judgement end, overwhelmingly so" because the institution was judged by their exam results and could not afford the luxury of letting 120 people engage in their own kind of professional development. Whether or not the manager would have publicly admitted to such an evaluative system remains unclear.

As at Rihab, some interviewees expressed a desire for a more developmental system. Amongst this group were two managers and five faculty members. This perspective was somewhat counter-balanced however, by three other faculty members, one of whom questioned whether the working week allowed any time for professional development, and two of whom questioned whether there was much more for such a well-qualified and experienced staff to learn.

Similarly, there was only one person who saw any problem with combining evaluative and developmental elements within the same system, though in contrast to Rihab, the lone voice at Al Fanar came from a faculty member rather than a manager. She began by stressing that appraisal should be separate from "evaluation and assessment ...[and] whether you are going to pass your probationary year". She continued by asserting that:

> I don't think you should be able to fail an appraisal. Somebody shouldn't come out of an appraisal session feeling, ah, I think I said the wrong thing. I think you should be able to say what you want, and your manager should be able to explore why you said certain things or why you feel certain things. But I think nothing that you say in your appraisal should count against you.

Very significantly, one faculty member took issue with the developmental / evaluative continuum and wanted to introduce a second dimension which she called the political / humanitarian because she thought "there are political purposes; there are manipulative purposes; there are organisational systemic things that don't fit on this [developmental / evaluative] line". During the interview, she drew figure 6.1 below. Her argument was that the appraisal system could be developmental or evaluative in terms of data collection, but the use to which that data was put could be humanitarian or political. So, in a heavily political setting, exactly the same data could be interpreted positively or negatively, depending upon whether management, for entirely different reasons, wanted to retain or reject a teacher.

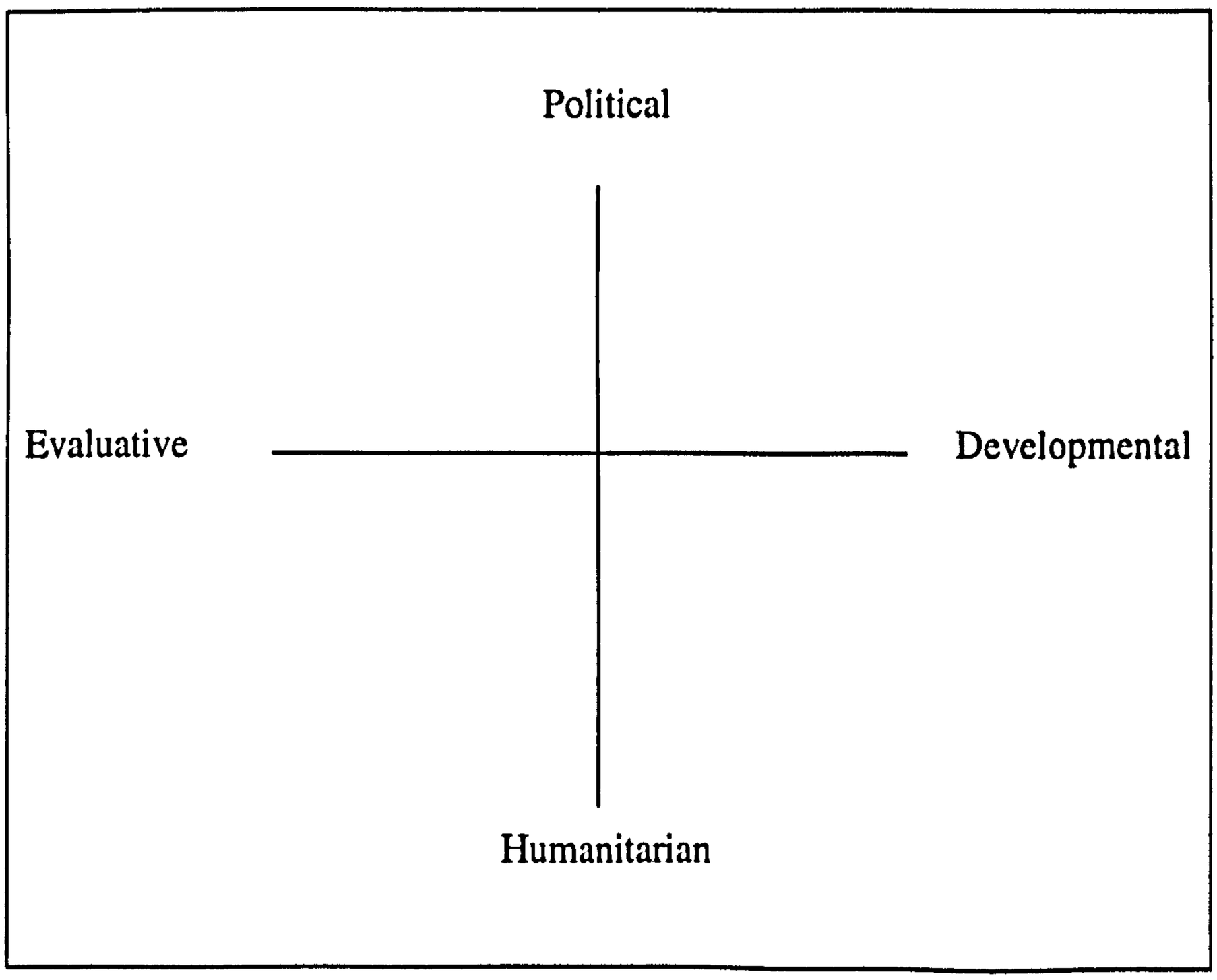

Figure 6.1

This "political" dimension of the appraisal system was also mentioned, at great length, by the faculty member who did not get her contract renewed, and by the person who 'failed' probation. More surprisingly, it was also

referred to by the senior manager. She described a person whose teaching was excellent, but whose public criticism of the director had cost her her job:

> On the teacher I'm talking about who left a few months ago, there was nothing in her file negative at all, at all. Not one word was there negative. Other people have been rehired for a three-year contract, and their file is that thick with complaints from managers and colleagues. But they get rehired or fired at the whim of the director.

None of the managers at Rihab made any similar statements, but, as discussed earlier, four faculty members did say the appraisal system might be used to gather evidence to support a dismissal decision made on other grounds. Moreover, a different Rihab faculty member made a less explicit reference to the kind of diagram shown above, when she suggested that:

> If people want to use them in a judgemental way, [appraisal] results will be interpreted how people wish to interpret them ...You can use a hammer to put a nail in the wall, or crack someone's head.

This "political" dimension to appraisal, explicitly mentioned at Al Fanar, and implicitly refered to at Rihab, does not figure prominently in the literature, but may well be the key to explaining why many teachers at both places seemed able to accept evaluative appraisal, in principle, whilst remaining deeply dissatisfied with the particular process they had witnessed at their own institution. Teachers at both sites reported that they were not averse to being judged, nor to seeing management remove people judged to be poor teachers. But they were averse to seeing colleagues they said were good teachers, and even "good team-players", removed for other reasons. This was mentioned specifically by nine faculty members at Al Fanar. The comments from three of them are quoted below:

> I could do excellent lessons from now until my end of my contract, but if there was a reason to get rid of me, I don't

think the lesson observations would make any difference whatsoever.

There is also a sense that appraisal operates on the basis of reaction to your opinions and your ideas, and not simply to how well you do your job.

I find it morally objectionable that teachers can have a good track record and be laid off without good cause, and often not told the reason, or told they don't fit. You know, it's just to me morally wrong.

Somewhat surprisingly, in view of the above, many faculty members still said the predominantly evaluative, and somewhat political appraisal system at Al Fanar was an improvement on the appraisal they had experienced in other organisations. Two people reported that they had not had any kind of appraisal in previous institutions and considered this a failing. Two other people with wide experience in the Middle East said the system at Al Fanar was more professional and more thorough than those they had experienced elsewhere. Similarly, one of the managers said she had "always experienced far less well-structured appraisal systems", both elsewhere in the Middle East and in the UK. She also spoke rather proudly of the fact that a colleague who was a headteacher in Britain had borrowed elements of the Al Fanar system for use in her own school.

In total, five faculty members and two managers said the Al Fanar system was better than any system they had experienced elsewhere, even though three of them expressed concern that in the year the research was conducted, some seemingly good teachers had not passed probation. By contrast, five faculty members and one manager said the Al Fanar system was worse than their previous appraisal systems. Two faculty members and one manager expressed a preference for their previous system because it had been more developmental, although both faculty members noted that the institutions in question had been much smaller than Al Fanar. The other

three faculty members said they preferred their previous systems because they thought the Al Fanar system was used unfairly. Not surprisingly, the person who 'failed' probation and the person who did not get her contract renewed accounted for two out of the three people in this group.

6.6.3 Comparison with the literature

Again, these findings tend to agree with studies of state school teaching more than studies of tertiary institutions. As we saw in the literature review, many commentators point to the apparent contradiction between an appraisal system that is simultaneously evaluative and developmental, whilst many empirical studies, such as those by Peaker (1986), Fitzgerald (2001) and Middlewood (2001), as previously discussed, report no such tension in the eyes of school teachers. In this respect, the findings of the current research concur with these school-based studies.

They also concur, in part, with Currie and Vidovich's (2000:143) study of Australian university faculty, where it is reported that "Many respondents made the point that they were not opposed to accountability per se, but to the form of the current requirements". Certainly, interviewees at both Rihab and Al Fanar reported no objection to evaluative appraisal, in principle, whilst, at the same time, around half of them (though not perhaps "many" of them) spoke out strongly against aspects of the specific system used at their institution.

Where the current findings diverge from previous HE studies is in relation to the apparent conflict between the educational values of management and faculty. This was discussed in considerable detail in section 6.1, where the findings of Randle and Brady (1997a and 1997b), Elliott and Crossley (1997) and Deem (2000) were examined. The dichotomy between management and faculty perceptions, reported by these writers, finds very limited support in my own research, either with regard to appraisal, in general, or with regard to the specific systems in place at Rihab and Al Fanar.

At both institutions, managers and staff agreed that the same appraisal system could be simultaneously evaluative and developmental; and at Rihab, management and staff disagreed equally amongst themselves over whether the current system was better or worse than what had been experienced elsewhere. There was some disagreement between management and faculty over whether the specific system in use was predominantly evaluative or developmental, with most faculty saying it was evaluative, and about half the management saying it was both evaluative and developmental. However, given the fact that only five managers were interviewed at Rihab and only four interviewed at Al Fanar, this difference is hardly indicative of a clear management / faculty divide.

There was also some disagreement between faculty and management over how far the specific system at each institution was negatively skewed, or open to misuse. At both places, around half the faculty made very negative comments about elements of the system, whereas none of the management were quite so openly critical. Here again, though, it is not clear how far this finding reflects a genuine difference of opinion, between the management and staff, and how far it is merely indicative of differing levels of interviewee candour. I suspect the latter scenario is more likely.

6.7 To what extent do teachers suggest they feel under surveillance?

6.7.1 Findings from Rihab

Not surprisingly, teachers were not specifically asked if they felt under surveillance. Rather, all of them were asked to comment on their experience of the student evaluation of teaching (SET) questionnaire, and on whether the appraisal system had had any effect on their level of stress. Of the fourteen faculty members, nine said the appraisal system had had no effect on their level of stress, although three mentioned that they had found other aspects of university life (such as their classes or managing their workload) stressful. One of the remaining five teachers said:

> Writing the [teaching] dossier adds one more thing to my list of things to do. In that sense, on a short-term basis, it does get me stressed out, because I don't have time to waste in such piddling little details.

Similarly, two other teachers mentioned being stressed when they received the results of their SET questionnaires:

> The student evaluation caused stress when I got the results. When I looked at it, there were really nice comments, which I sort of passed over really quickly, and then there were very negative comments, which I dwelt on for days.

> I received a bad evaluation from one class and that was quite stressful, but I went to my line-manager who said, 'Don't worry about it'; she got a worse one.

The remaining two teachers found the appraisal system extremely stressful for reasons they explained at length and with visible emotion. In one case, the teacher related how her Dean had tried to get her fired on the basis of a single set of poor evaluations from some well-connected students. She added, however, that the Dean may have been using this as an excuse to get rid of her because she had been heavily involved in a curriculum proposal that the Dean had not taken up, but which was attracting favourable attention from more senior management. On the morning that she had been invited to discuss her proposal with one of the Vice-Chancellor's advisors, she related how:

> The morning in question [the Dean] was waiting for me in [my manager's] office ... and basically threatened to fire me on the basis that my student evaluations weren't good ... She had never observed my class, and neither had [my manager]. They had never set foot inside of one of my classes. My evaluations were high [the previous semester] ... and without having observed me, and on the basis of no other information than the student evaluations, and the fact

> that [an advisor to the Vice-Chancellor] wanted to see me about the curriculum proposal, she said, quote unquote, that I must be neglecting my classes because I had time to write proposals. And therefore, that explained the negative student evaluations.

The other teacher had had a similarly negative experience. She said that the appraisal system had had an "extreme" effect on her level of stress in the semester in which the interview was conducted. She had been "totally paranoid ... that anything I did would be taken negatively". This was the result of a poor SET rating ("some of the comments ...were quite positive but the numbers were extremely low") and a "relentlessly negative" post-observation feedback session with a manager. Clearly fighting back tears, she explained how:

> It was an extremely negative experience because the person did not have anything good to say. It was all relentlessly negative. And I left in tears, because it was so bad. And I don't think I'm a bad teacher and I don't think it was a fair evaluation. I don't think it was called for.

The teacher thought she had received a low student rating partly because some of her students had received low marks and partly because the student questionnaire did not distinguish between the teacher and the course. This opinion was repeated several times in the interview and the extract below represents just one instance:

> The way the current document is, it's not a clear evaluation of the teacher, it does not separate teacher from course. So, if the course is poorly designed which is not something the teachers have much control over, the teachers get marked down.

Clearly, these two teachers had very negative appraisal experiences, and the strength of their feelings cannot be ignored. Nonetheless, they represent

only two of the fourteen respondents, the majority of whom did not appear to find the appraisal system stressful, per se.

Opinion was more divided over whether the appraisal system might be used by the management to dismiss faculty. Although several interviewees said that in theory the appraisal system should highlight people not doing a good job, no-one gave any specific examples of this happening in practice. Indeed, two faculty members said they thought it quite unlikely that anyone would suffer any kind of censure as a result of the appraisal system. One said that in order not to pass probation, a faculty member would have:

> To be doing something really silly, I think, something to be noticed in other ways, 'cos I don't think the [appraisal] instruments will necessarily show that.

Similarly, the other said:

> I don't really feel like I'm going to be fired. And I don't feel like they're likely to fire anybody really over the whole dossier evaluation thing. Maybe if there were other reasons, combined with bad student evaluations, but I don't think bad student evaluations would be cause for them to fire anybody, at this point. So, I don't feel like my job depends on it.

In addition, two other faculty members recalled how management had ignored consistently low SET marks, showing that the results of the appraisal system were not being acted upon. In the words of one of them:

> People sail through their probationary period, even though they get trashed by every student they teach.

In a field note, I recorded exactly the same point being made by two of my own students, a few months earlier. They had come to get their exam results and mentioned that despite getting 'A' grades in most of their subjects, they had failed one particular subject, taught by a "hopeless" person their friends had complained about the previous semester. Most

pointedly, one of them asked if anyone ever read the student evaluations, because nothing had been done about it.

If the comments of these students and teachers are to be believed, there is a strong suggestion that the appraisal system is seen as ineffective, because very often it does not identify poor teaching, and without exception it does nothing to correct it. This does not mean, however, that faculty think it difficult to lose their jobs. On the contrary, two managers and four faculty members mentioned how they felt a teacher could quite easily be dismissed, but for reasons that were not related to the formal appraisal system. One of the managers, who had worked in the country for many years at a variety of federal HE institutions, commented:

> Occasionally there may be problems in the classroom – the style of teaching just does not fit with the philosophy of the teacher, the philosophy of the organisation or the institution, but most of the time, when people are asked to leave, or when the individual is not renewed, it's because of some interpersonal problems that they're having.

Somewhat surprisingly, all four of the faculty members who expressed a similar opinion went on to say that they were not worried by the fact that they could be dismissed at any time for any reason. Three of them commented as follows:

> I mean the kind of culture that it is here – people can get sent home any day for any reason. The appraisal system would give one piece of evidence, but they could do anything - they could send us home for traffic fines. So I'm not concerned about that. It's no worse than any other particular piece of paper we generate here ... Last year, when I had a particularly stressful semester teaching the repeaters, I came to the realisation I don't care. (Laughter) ... If they put me on the plane, I'll go back home, and I won't be missing anything here. So, I will teach here as long as I'm welcome, and when they're tired of me, I'll go home.

I'm not worried about it ... but the simple fact of it is that you can lose your job here for more reasons than professional competence, and you may never know what those reasons are and it's not going to be any different the day after I pass probation than it was the day before.

Faculty, staff and administrators routinely leave or they are fired, and it doesn't seem to have any relationship whatsoever to faculty appraisals by students or anyone else ... I think that the hiring, rehiring, and firing processes from the administration's standpoint are so capricious that they really don't matter that much. I know that I'm not doing anything routinely that is contrary to the university policy in the classroom, and so I'm not worried about my evaluations leading to my getting fired.

In addition to these four faculty members, there were another four teachers who suggested that the appraisal system could be (mis)used as 'an excuse' in cases where management wanted a person dismissed for reasons unrelated to the appraisal system criteria. The first of these faculty members had had the extremely negative experience with her Dean outlined above. She suggested that one management benefit of the system was that it provided them with "an endless supply of material to use to get rid of somebody, if someone has been an annoyance in some way, even unintentionally".

The other three faculty members focused specifically on how the results of the student evaluation might be used as an excuse to dismiss someone who had attracted unfavourable attention for other reasons. One of them commented:

It [the SET questionnaire] is something that if they want to, the administration could possibly use against you, if they want to ... I've a feeling that if, for any reason, they didn't wish to renew somebody's contract, then they could look

> rather closely at this, and find possible reasons based on student evaluations.

This perception by faculty that anyone could be dismissed at any time was also mentioned by three of the managers, although one of them thought the university was not unique in this respect, and another thought this perception might be a distortion of reality. The strongest statement came from a manager who said:

> I think this is a very uncertain place to work in. I have never been in a place where I think there is so much uncertainty ... it [overnight dismissal] can happen and everybody knows an example of when it did, and so therefore it just makes everybody uneasy.

Two other managers were more muted in their comments. One said:

> There is nothing unique about the feelings engendered by the system at Rihab, although here it can be more extreme, quicker, and more brutal in that administrators, but not faculty, have been dismissed overnight. A high percentage of the fear amongst faculty may be self-manufactured, but understandable in the context. No faculty member in English has been dismissed because they got a bad evaluation.

The other said:

> There's a sense here that nobody is ever safe. Those who've completed their probation, some of them are still worried.

When asked to elaborate on what she meant by "nobody is ever safe", she continued:

> I said that's a perception. I wouldn't say it was necessarily true ... [It came about] ... because there have been so many sackings in the last year. But not, to be honest, sackings of faculty, very few faculty here have been fired.

On the basis of these data, I think it is fair to say that most faculty members at Rihab do not feel directly under surveillance as a result of the appraisal system, in general, or student evaluation, in particular. Many do believe that they could be dismissed at short notice, for a wide variety of reasons unrelated to their teaching, but most do not seem worried by this.

6.7.2 Findings from Al Fanar

As at Rihab, faculty at Al Fanar were asked to comment on their experience of the student evaluation of the learning environment (SELE), and on whether the appraisal system had had any effect on their level of stress. With regard to stress, the results were quite similar. Twelve of the faculty said the appraisal system had had very little, if any, effect on their level of stress. Some of the more long-standing teachers remembered it being more stressful in their probationary year, but five probationers were actually included in this group of twelve. By contrast, three other people described the process as "incredibly" or "extremely" stressful. Of these three, one teacher was in the first year of her second three-year term; one had just passed probation; and one had sought and been denied a third three-year contract. As at Rihab, these three people spoke at length and with obvious emotion about their experience. Notwithstanding this, their perspective did not seem to be shared by the other twelve faculty.

Opinion was much more divided over the importance management attached to the SELE results. As at Rihab, the managers themselves tended to downplay its significance. One said it could be "very, very informative", but it could also be "judgemental", and was "not in any way indicative of whether a teacher is doing a good job or not". According to my field notes, another told a meeting of faculty that they looked for "themes" and "patterns" in the SELE results rather than making decisions on the basis of one set of results. She reassured faculty that they took account of the type of students each teacher was working with, and told an anecdote about teaching a class of multiple repeaters in her first year at the college and having her manager tell her that if they did *not* give her a bad evaluation, she was obviously not doing her job properly. A third manager said student

"feedback" (the word she preferred over evaluation) was "useful" but less important than her own observations. The fourth manager reported that student evaluations would only be used to confirm what she already thought about a teacher. Bad evaluations about a teacher she thought was good would be ignored; bad evaluations about a teacher she thought was not doing a good job would be seen as agreeing with the "expert and professional". At a faculty meeting, she called it "one tool of many" and said "no-one had ever lost their job on the basis of student feedback alone".

Interestingly, despite what the managers told me in their confidential interviews, and, more importantly, what they said publicly in faculty meetings, only a minority of staff appeared to believe them. Four people reported that, in their view, management did not attach much importance to the results, unless they were consistently low. One of them recalled how a manager had become concerned after students had complained a number of times about a particular teacher and had therefore done a couple of extra lesson observations. In her words, the manager had described the second observation as "a perfect lesson" and "that was the end of the story. These students kept complaining about her [the teacher] but she [the manager] was absolutely convinced it was the class's problem not the teacher's".

On the other hand, seven faculty stated that a great deal of attention was paid by management to SELE results, and that student opinion, however it was expressed, carried a great deal of weight. One put it most succinctly by saying "if a student complains, they could sack me", whilst another described two occasions in the last five years where teachers she thought were competent had been attacked by students behaving "like a lynch mob" with the result that "those people lost their jobs".

Another faculty member mentioned how student evaluation was used elsewhere to improve teaching, but at Al Fanar, it was "used to judge a teacher and to piece together evidence in order to decide who to keep and who to let go". In other words, SELE results formed the basis of contract

decisions. By contrast, a different faculty member suggested that SELE results were used to support contract decisions made on other grounds:

> Look at the way the student complaints are dealt with. I could name three members of staff, serious student complaints made against them, a series over the last four years, and nothing has been done about these complaints. They are made; they are routinely filed. On the other hand, if there is a member of staff whom management wishes to eliminate, usually for political reasons or because that person is a perceived dissident, then those student complaints don't go into the file, they become the basis of disciplinary action.

Wherever the truth of the matter might lie, it seems clear from the data cited above that, despite management reassurances to the contrary, the student was indeed seen by many Al Fanar faculty as "a surrogate surveillance device" (Randle and Brady, 1997b:133).

6.7.3 Comparison with the literature

Cullen (1997:189) writes of a concern amongst commentators "that appraisal would be used as a means of surveilling and controlling educator behaviour". The headteachers in her study seem to have overcome this concern and speak very positively of the experience. Their reported enthusiasm for the process and their professed belief that appraisal was not being (mis)used in this way finds considerable resonance with the views expressed by the management in my own study.

With regard to faculty, however, the picture is rather different. We saw earlier, in the section on shifting power relationships, how, according to Randle and Brady (1997b:132-133), "notions of common enterprise, co-operation and mutual responsibilities" have been usurped by "surrogate surveillance", and how a formal complaints procedure may seem like "a reasonable and democratic mechanism designed to protect the students",

but is, in fact, "a potentially destructive instrument that could undermine the authority of lecturers in the perception of the student".

It is unclear whether the faculty in my own research ever enjoyed the "traditional" kind of lecturer / student relationship described by Randle and Brady (1997b:132), but undoubtedly, many of them were now very concerned about the amount of power apparently wielded by the student population. Randle and Brady (1997a:238) attribute this development to the process of "marketisation", but, in the country in which I conducted my research, all public HE is free for nationals, and therefore, other factors, relating to the specific social, political, economic, and cultural context, seem to be a more likely cause. When discussing appraisal in general, or the power of the students, in particular, respondents from Al Fanar made reference to the fact that the great majority of the population are 'temporary expatriate residents' and only a very small minority are nationals, whilst respondents from both Rihab and Al Fanar made reference to the fact that the employment market is very tightly controlled by the concept of 'sponsorship'. These two factors make the situation in the country in question very different to that in Britain, North America, Australia or New Zealand, and an understanding of their influence is essential in determining how far the findings from this study are generalisable to other contexts. For this reason, they are explained in more detail below.

6.8 The wider context

6.8.1 Demographics and the power of the student

Rihab and Al Fanar are located in a Middle Eastern country where only a minority of residents are indigenous citizens having an unlimited right of abode. Exact figures are hard to come by but one source (The Economist Intelligence Unit, 2001) estimates that 20% of the population fall into this category, while another source (The World Factbook, 2001) puts the figure at 35%. The remaining 65-80% of the population are foreign 'guest workers' and their dependents, who have been invited to work in the

country by the private or public organisation that sponsors them. At the time of the research, foreign guest workers were not allowed to become permanent residents, even though some of them had lived and worked in the country for several decades.

As has been mentioned previously, only a very few faculty at Rihab said students had any power to influence a teacher's future prospects at the institution. By contrast, just under half of the faculty at Al Fanar suggested students, on occasion, had the power to get teachers dismissed. One faculty member said she thought the situation here and in the States was similar in that "If you run afoul of your students, you could lose your job both there and here". Several other faculty members, however, suggested that students here had considerably more power than they would have elsewhere. One faculty member remarked how:

> Students have a lot of power here, I think. And a lot of teachers are afraid of that ... in Western countries, if a gang of students went and complained to a head or a director, about a certain teacher, and say, for example, that head fired that teacher, well, there's recourse, there are legal ramifications, and there are ways to see justice done ... but here, we don't have that security, so that's where the fear comes in.

When asked why she thought the situation here was different to that in Western countries, she replied:

> There's a culture here of 'we own this land, we are the citizens, we are the locals, and therefore, we get anything we want'. And basically that attitude, even though they might not voice it that way, but basically you see that attitude in so many things, in their daily life. 'We deserve anything we want', they have that kind of feeling. So if things aren't the way they want it around the college, they feel they have the right and the power to do something about it, and actually they do.

Similarly, another faculty member "speculated" (her word) that students with *wasta* (meaning socio-political connections) "could" exert an influence not normally encountered elsewhere in the West, though she did not know for certain that they did. Her exact words were:

> But in terms of informal appraisal of teachers, then, wasta is going to have an influence. And you don't encounter that usually, in other educational locales ... I mean that it would be unusual, for instance, in Western countries for a student to have a lot of pull in terms of their opinion of a teacher.

A third faculty member told an anecdote about visiting a student on work placement at a police station and watching her order around her expatriate boss and click her fingers to have tea brought in. She remarked that:

> Elsewhere the people who are on work placement ... are the lowest of the low; they are below the lowest employee ... But here, our students go out and they are absolutely nowhere near the bottom of the pecking order.

Even one of the managers talked about expatriates being "allowed and tolerated for a specific purpose" and being asked to leave if they did not fulfil this purpose. She added that "teachers are viewed as a commodity, for want of a better way of expressing it".

Now it may well be the case that teachers in other contexts are also viewed as a commodity, but the employment laws of the country in question, and particularly the concept of sponsorship support this perspective to an extent that many respondents thought was not replicated elsewhere.

6.8.2 Sponsorship and the lack of legal constraints

As mentioned earlier, people who wish to work in the country in question need to find a public or private organisation that will sponsor them. The sponsor arranges an employment visa for each worker and, according to the labour law, the sponsor has the right to cancel this visa at any time for any reason, although, in certain circumstances, he may be required to pay the

employee three months salary, by way of compensation. Once the guest worker's visa has been cancelled, she must leave the country within 30 days. The sponsor is also able to enter a "ban" in the guest worker's passport, which prevents her from being employed / sponsored by any other company in the same country for a period of between six months and one year. (Further details about the exact provisions of the labour law of the country in question can be found at http:///www.tamini.com/publications/labourd1.htm)

Employees sponsored by the federal government are actually exempted from the labour law described above, but the provisions of their contracts are broadly similar, except that they must be given either six months notice of termination for non-disciplinary reasons, or six months salary in lieu. Federal employers do not usually impose a ban on teachers, but very few of them ever secure alternative employment within the 30-day grace period.

Many of my informants referred to this concept of sponsorship, either directly or indirectly, and it clearly has a major impact on how appraisal is perceived here. To be fair, there was one person at Rihab who said, "if somebody's doing their job, even if it's not up to whatever standards they've set, then they're probably quite prepared to go along with it, if the person's not rocking the boat too much". However, there were also six other teachers who discussed the ease with which they perceived people could be fired, although three of them thought this applied more to managers than teachers. Their comments included the following:

> In this system, it's not like the States – you don't have recourse to say 'I have tenure'. Or, you don't have recourse to say, you know, you don't just get fired on whims, you know ... [here] I think people who get fired because of a personal situation ... there's not much they can do, there's not much that they can say.

I think the notion that we're on probation for one year is farcical, because we're always on probation in the sense that we can have our contracts terminated any time ...

I think it's quite easy to get sacked here.

There haven't been that many teachers that have been fired, that I know of. There have been a lot more administrators fired ...those people come and go like the wind.

> Look at all the heads that rolled [in the] Spring term ...I mean the reign of terror was at [Rihab]. Every day you came in, who else's head had rolled?

With regard to managers, as mentioned previously, one of them said that the feelings engendered by the system at Rihab were not unique, even though the overnight dismissal of administrators was more extreme and more brutal than elsewhere. By contrast, a different manager commented on how very dissimilar she thought Rihab was to other places she had worked:

> This is the first time in which I've ever been where I couldn't protect the faculty ...And I just don't like that feeling. Not that it's paternal. I don't mean it in a paternal way, I mean it in the way of management, in the sense that they can't come and dismiss you without my being involved and either agreeing or trying to defend you or whatever. There is no appeal process here, which is such a big thing in American universities now.

Perhaps the situation is best summed up by the third manager, quoted earlier, who said there was "a perception" amongst staff that "nobody is ever safe", even though she "wouldn't say it was necessarily true" because "very few faculty here have been fired".

Sentiment at Al Fanar seemed even stronger. Six faculty and two managers mentioned the lack of legal constraints. One person referred explicitly to the system of sponsorship, saying:

> In my opinion, the problem is the society. Or a set-up where your job and your right to live in the country, and your children to go to school, and have a house depends on your job. And if you lose your job, you've got no recourse. You can't go to any industrial tribunal to get it back. You have to hand over your house, the school, and leave the country in most cases, unless you find another [job]. And that employer can control whether you can move to another country or not ... It's because of the sponsorship system ... which I don't think is conducive to good management practices, or any appraisal system at all. I think that's what it boils down to.

Another talked about how management were "not constrained by any legal or procedural or rules or anything ... so therefore, the use of power is overwhelming". Likewise, the person who 'failed' probation suggested that the administration of the college:

> have a policy of letting people go because there are no constraints, as there would be in Western countries. Institutions like the employment court and so on. Here they don't have any constraints. So they can feel free quite legally and morally to sack people that they think can be replaced.

Somewhat surprisingly, one of the managers was equally forthright. She suggested that "each college is a little dictatorship ... one wrong word and we could all be out". When asked to elaborate on her choice of the word "dictatorship", she continued, "all power is in the hands of one person [the college Director] and nobody else can affect that decision ... and that wouldn't happen if we were in a country and a system that had checks and balances in place, but we aren't".

Chapter 7: Tentative Theory Generated

7.1 Summary of research questions and findings

7.1.1 How is appraisal, in general, perceived by faculty and managers? How far are accountability models acceptable to either group?

Faculty and managers at both institutions commented on the need for appraisal in academia, concurring with much of the literature. They did not voice any opposition to the principle of accountability, and, although perceptions differed on how to measure quality and how to balance accountability and professional autonomy, these differences were not split along partisan lines, providing little evidence of the management / faculty divide over educational values highlighted by previous research into HE management. (See Randle and Brady, *1997a and 1997b*; Elliott and Crossley, 1997; and Deem, 2002).

7.1.2 Is appraisal perceived to have any effect on the quality of student learning?

Only six out of 29 faculty members suggested that any aspect of the appraisal system at their institution had had any kind of positive effect on their classroom behaviour. This finding concurs with other empirical studies, such as those by Turner and Clift (1988), Kyriacou (1995), Wragg et al. (1996), and Cullen (1997), though it remains at odds with much of the earlier non-empirical rhetoric on appraisal, such as that by Powney (1991a and 1991b), Mortimore and Mortimore (1991), Fidler and Cooper (1992), and Magennis (1993).

7.1.3 What connections, if any, do teachers make between student evaluation of teachers and grade inflation?

Four out of 29 faculty members admitted inflating grades in an effort to secure favourable student ratings, as opposed to fourteen who categorically denied this, and eleven who made no comment either way. These findings again concur with much of the literature, particularly from the United

States, in which teachers, themselves, report very little evidence of student evaluation leading to grade inflation.

7.1.4 How much time do faculty and management devote to appraisal? Do they consider that the time devoted to appraisal is time well-spent?

Managers at both Rihab and Al Fanar commented on how time-consuming appraisal was, particularly for middle managers, a very common finding in the literature, examples of which include the work of Turner and Clift (1988), Kyriacou (1995), Cullen (1997), and Campbell (2002). At Rihab, there was some suggestion from some managers that the time was not particularly well-spent, but at Al Fanar, management spoke repeatedly of the importance of appraisal and the great need to find sufficient time for it. By contrast, most faculty members at both institutions said they spent only small amounts of time on appraisal, a finding somewhat at odds with the studies cited above. Nonetheless, four teachers at each institution still said appraisal was a waste of time, either because it had no discernible impact, or because the management at Rihab, in the first year, were alleged not to read the data they collected.

7.1.5 To what extent is appraisal said to focus on either basic technical competencies or flexible, creative practices?

Managers and staff at both institutions commented repeatedly on the need for appraisal to be fair; the need for evaluation to be objective; and the need for decision-making to be evidential. Most interviewees noted that the observed lesson should be a contrived performance, and some also suggested the class visit criteria focused on behaviour that was easily observed, but somewhat superficial. This was also said about the student evaluation of teaching questionnaire. To a large extent, therefore, appraisal was indeed reported to focus on basic technical competence. However, for the most part, this was considered a wise precaution, since the appraisal system at both places was said to be highly evaluative, and the student population somewhat whimsical.

7.1.6 How far is the specific appraisal system at each institution perceived as developmental and / or evaluative? How far is it seen as embodying a paradigm of professionalism and / or managerialism?

Almost without exception, faculty at both institutions said the appraisal system was predominantly evaluative. Three of the five managers at Rihab and two of the four managers at Al Fanar agreed. Most interviewees said they wished the current system were more developmental, and could see little problem with simultaneously combining evaluation and development. Such a finding supports previous empirical research, by Peaker (1986), Fitzgerald (2001) and Middlewood (2001), in which teachers were reported to want an appraisal system that was both evaluative and developmental.

Moreover, several teachers, particularly at Rihab, said they were not adverse to being judged, but did not think the specific appraisal scheme currently in use at their institution was a fair instrument with which to assess their performance. In other words, an evaluative system that was perceived to be fair would not necessarily be considered an imposition of managerial control. Fairness was the key factor, not evaluation, per se.

7.1.7 To what extent do teachers suggest they feel under surveillance?

Many faculty members, particularly at Al Fanar, spoke of the unusual amount of power seemingly wielded by some students, and the ease with which people on both sides of the management divide could lose their jobs, sometimes because of their appraisal results, but most often in spite of them. Generally, this was attributed, not to the increasing marketisation of education, as is the case elsewhere in the West, but to the specific social, political, economic, and cultural context of this particular country, where up to 80% of the population are foreigners, subject to sponsorship laws that have no parallel outside the Middle East.

7.2 Fuzzy generalisations and worthy conundrums

This section will focus on the tentative conclusions and "fuzzy generalizations" (Bassey, 1999:62) I believe can be drawn from the data

discussed above, and point out some corresponding conundrums I believe are worthy of further investigation.

7.2.1 Appraisal is said to have very little effect on classroom practice. Why should this be the case?

Despite the considerable amount of time that managers devoted to appraisal, only six out of 29 faculty members said there had been any resulting change in their classroom behaviour. Part of the reason for this might be that many of the teachers at Rihab and Al Fanar were highly qualified and experienced classroom practitioners, a conclusion that has implications for the concept of continuing professional development. And part of the reason might also be that feedback was generally numerical and lacking in specific details.

This suggests that the extent to which appraisal results in improved teaching depends on how qualified, experienced, reflective and open the appraisee is; how well-respected, sensitive and skilful the appraiser is; and how detailed and context-specific the feedback is. It also suggests that over-emphasising the classroom-based consequences of appraisal may lead to a regrettable neglect of more intangible, but equally significant, effects on the professional identity of individual teachers, and the overall management of the academic institution. The impact of appraisal on these two areas obviously warrants further study.

7.2.2 Appraisal is said to embrace the deficit model of teaching as either a labour or a craft. Is this inevitable?

According to House and Lapan (1989:56), teaching can be seen as a labour, a craft, a profession or an art. Where teaching is considered a labour, "then the teachers' job is to implement preset, prescribed procedures and routines, and the appraisal system includes direct inspection of the teacher's work, such as the monitoring of lesson plans, classroom performance and performance results" (House and Lapan, 1989:56). Sticking to procedures is paramount and close supervision is required. If teaching is seen as a craft, then evaluation determines whether the teacher has the appropriate

"repertoire of specialised techniques and knowledge, including knowledge of how to apply these skills" (House and Lapan, 1989:56). By contrast, teaching as a profession "requires not only the repertoire of skills but also the exercise of judgement about where the skills should be applied," whilst teaching as an art requires that:

> the techniques and procedures of teaching are personalised rather than standardised because the teaching situation is seen as unpredictable, requiring a frequent departure from set rules and techniques, and these departures are an expression of the personality and personal insight of the individual teacher.
>
> (House and Lapan, 1989:56)

According to the interview data, appraisal at both Rihab and Al Fanar was seen by almost all faculty and half the management as evaluative, predicated on a labour / craft model of teaching, and concerned with the maintenance of minimum standards of competency. It was not seen as an opportunity to engage in collaborative experimentation and critical reflection.

Moreover, two managers at Al Fanar commented on the fact that identifying any weakness on the part of a teacher could be their death-knell. The senior manager talked of putting together a very successful action plan which helped turn around a probationary teacher, only to have the most senior member of the college dismiss her anyway. In her words:

> identifying them as being at risk has more-or-less been their death-knell ... whatever they have done has not been enough to satisfy the management that it's good. And this upset me last year, because there was one person ... we'd put together this plan for her which included lots of peer observation, lots of lesson plans, lots of discussions. But, I'm afraid this was almost seen as a kind of weakness. And I was most upset that we lost that lady, because she did turn around. She wasn't at the beginning, but by the time April / May came

> round, she was a fine member of faculty. But unfortunately the decision had already been taken long before the plan had come to fruition ...

The other middle manager was more succinct, simply saying:

> One of the problems might be, instead of allowing time for problems with one particular teacher to work themselves out, PD, support, help and so on - depending again on who's up at the top, maybe that stops being the case. So, the minute anything, any problem is mentioned at appraisal, it's bye bye.

Similarly, as noted before, one of the appraisers at Rihab admitted giving top marks to everyone, during the class visit, because anything else "could so easily be misinterpreted" at a later stage. Judging by the fact that all but one of the teachers at Rihab who discussed their numerical rating said they got top marks in all categories of the lesson observation form, this appraiser was certainly not the only one to adopt such a strategy.

All of which suggests that when appraisal operates on the basis of the deficit model of teaching, this is not so much inevitable, as the consequence of a particular organisational context. As one interviewee put it, to ensure students get the best possible teaching, you need an appraisal process that is honest and open; and to have that, you need an atmosphere of trust. "So, it's a chicken and egg sort of thing."

Thus, it seems fair to conclude that the more evaluative the appraisal system, and the more punitive the overall organisational culture, the greater the concentration on basic competency, the greater the lack of constructive feedback, and, paradoxically, the greater the potential for appraisers 'turning a blind eye', at least on paper, to inadequate performance.

7.2.3 Appraisal is said to bear very little relation to continuing professional development. What does this tell us about the concept of CPD?

At each institution, only one faculty member suggested that the current appraisal system had been in any way developmental, though many others said they would have welcomed the opportunity to act on more detailed feedback. Appraisal, therefore, not only has little impact on classroom performance, but it also has little impact on continuing professional development. Clearly this is related to the fact that appraisal at both institutions was seen as operating on the basis of a deficit model of teaching, with very little emphasis on development. Indeed, a faculty member at Rihab and a manager at Al Fanar both said they believed there would be little effort to improve the performance of people deemed not to be working at the right level; they would simply be asked to leave. The comments of the manager at Al Fanar are particularly apposite in this respect, and worth quoting in full:

> I have to say, I don't think that the organisation is set up in such a way that it can be particularly formative because we don't really have the environment in which you can work with somebody and develop them. So, to a certain extent, we are saying to people, this is the level of performance we expect, right from the beginning and through our appraisal system, we are looking to see if you meet that standard, and if you don't, there's not an awful lot we can do about it, really ... I think it's political basically, They pay good salaries and they can afford to - they say, okay, you know, we'll give you this, but this is what we expect in return. And we're not really interested in having to develop people because we can get other people who are as good or better. I think it's simple economics ... I have mixed feelings about that. Because in previous organisations I've worked for, it's not been like that, and we have worked with people. And I think, you know, it's obviously much more preferable to work with people and develop them. The problem is, that

> while you're dong that, students can be suffering. And I think that's fair enough, you know, if it might take somebody a year to come up to an acceptable standard of teaching, and in the meantime, the students could be getting more from another teacher who's already there.

Likewise, another of the managers at Al Fanar suggested that the institution needed better than average teachers because the students had had such previously poor teaching in government schools. In her words:

> There will always be good teachers and bad teachers, and there will always be average teachers. And the stakes here are higher because our learners have so many difficulties and because they are not independent learners and because they are so immature. Because their experiences have been so paltry in education. So the stakes are higher. We need better teachers, it's as simple as that. Students in other areas, or other countries that could get away with an average or even a mediocre teacher, it just doesn't happen here. They have to have the best, if we're going to succeed.

The implication that certain institutions have the right to employ only the very best teachers has huge ramifications for the profession, and challenges the very concept of CPD. On the one hand, it has become something of a mantra in education that all teachers should engage in life-long learning. On the other, it goes almost without saying that any educational institution that makes teaching its highest priority will seek to employ the best possible classroom practitioners.

How these two apparently contradictory ideals might be reconciled is far from obvious. Clearly, it is beyond the scope of my current research to explore the concept of CPD in any great detail, but the following questions are certainly worthy of further investigation:

Does the exhortation to improve in the future necessarily imply a current weakness? When we talk of CPD, do we mean learning to do the same job better (such as classroom teaching), or do we mean learning a slightly different job (such as testing, management, materials writing, and so on)? If we mean learning to do the same job better, what does this tell us about the impact inexperienced teachers have on student learning? What does this suggest about an institution's responsibility to the profession as a whole, as distinct from their responsibility to the specific individuals enrolled in their classes? Should certain institutions, by virtue of their financial bargaining power, or allegedly disadvantaged student populations, be able to 'cherry pick' the best teachers?

None of these questions has an obvious answer, but the striking suggestion that students may suffer while teachers are being developed ought not to be ignored.

7.2.4 Appraisal is said to function, at times, as a mechanism to further micro-political machinations. How is this possible?

Four faculty members at Rihab noted how the appraisal system was ineffective in identifying poor performance; another four faculty members and two managers said it was easy to get the sack, but for reasons unconnected with the appraisal system; finally, a different four faculty members mentioned how the appraisal system could be misused to provide evidence in support of employment decisions made on other grounds. In this way, it was reported that appraisal results could be used selectively by management to support their actions, but not by faculty to contest them.

Similarly, as was discussed in section 6.6.2, nine faculty members at Al Fanar made reference to the fact that people they believed were good teachers, if not good team-players, had nonetheless been asked to leave because, in their interpretation of events, management thought they did not fit in, or were not sufficiently compliant. In the year that my Al Fanar research was conducted, three people did not pass probation, and, interestingly, both the appraiser concerned and the senior manager said, in

their interviews, that these particular decisions were entirely justified, because the probationers' teaching had not been up-to-scratch; in contrast, none of the faculty interviewees expressed the same sentiment, and, indeed, the majority of them said the exact opposite - that it was blatantly unfair because their lesson observation scores had been no worse than anyone else's.

Now, of course, this could be an inevitable 'blind spot', in as much as faculty will agree, in theory, that incompetent teachers should be dismissed, but never actually recognise such an eventuality in practice. This seems unlikely, however, for two reasons. Firstly, two Al Fanar interviewees gave specific examples of teachers at other colleges being let go, for what they said were justifiable reasons; secondly, one Rihab interviewee said she knew of several teachers who should have been dismissed and were not. In view of this, it seems fair to conclude that teachers are not averse to evaluative appraisal, nor to seeing poorly performing colleagues asked to leave. What they object to is teachers whom they see as good classroom practitioners being asked to leave for reasons they either do not know about or do not agree with.

Obviously, it is impossible to know 'the truth' of the matter in this instance; trying to discover how good or bad the probationers' teaching actually was would be naïve in the extreme. What is significant is the conflict of opinion between management and faculty, and the effect on staff morale of the perception that management, in the absence of any legal constraints, are able, at times, to abuse the appraisal system to fulfil micro-political ends.

Undoubtedly, a certain amount of antagonism between management and faculty is evident in very many institutions around the world. One of the Al Fanar interviewees suggested, only half jokingly, that "it's your job really, to make fun of, or find fault in management, not being one myself". Likewise Turner and Clift (1988:64) contend that:

> The notion of the senior management team as a sophisticated, conspiratorial group with dictatorial aims and

> Machiavellian means is one that some teachers in junior or middle management positions quite readily espouse, but one wonders whether in many schools senior management is able to adopt so pro-active a stance.

In this respect, Rihab and Al Fanar are very similar to thousands of HE institutions worldwide. What makes them different, however, is the extent to which legal constraints on the misuse of power are absent. This was explored extensively in section 6.8 and need not be repeated here. Suffice to say, management and faculty at both institutions recognised how easily the appraisal system *could* be abused in situations where staff were not eligible for tenure; trade unions were non-existent; and sponsorship laws were heavily weighted in favour of the employer.

7.2.5 HE in the UK, the USA, Australia and New Zealand is quite different to HE in the Middle East, in terms of the social, political, economic, and cultural context. What, if anything, can the former learn from the latter?

As has been previously mentioned, many interviewees commented on how different the context of HE in the Middle East is, when compared to other places. A manager from Rihab, and a faculty member from Al Fanar both noted how difficult it was in North America to remove people who persistently under-performed. One of them told an anecdote about how her spouse, a senior academic at an American university, had tried to dismiss a member of faculty who was "out of control", and then been sued for sexual discrimination in a suit that "dragged on for years". She suggested "it is literally impossible to dismiss anybody in the universities now in the States. Now that's going a little bit too far". Likewise, the other interviewee described her experience in Canada, where "in a union environment, appraisal becomes fairly meaningless because it's almost impossible to get rid of anybody ... [even] ... horrible, completely lazy, horrible, horrible, useless workers". Both these people contended that it was a good thing for management to be able to dismiss people whose work continued to be substandard, despite repeated opportunities for professional development.

The crucial point, however, was that such freedom must be exercised with extreme caution and complete fairness. In other words:

> The system here of being able to get rid of people is good, but being able to get rid of people for no reason without criteria, not explaining, not knowing, the people themselves don't even know why they are fired, that's not okay, that's going too far ... the system they've got here could be great ... because they've got the power to do a great job, but they have to take that power and turn it into an altruistic humanitarian way, and not just for the purpose of amassing their own power and maintaining that hierarchy.

Obviously, most people who are asked to leave their place of employment do not publicly admit that their performance was inadequate, or even that their profile was unsuited to that particular educational context. It is unrealistic to expect people who are sacked to agree openly with the decision. On the other hand, it is vital for the sake of staff morale that any such decision is seen by the rest of the staff as fair, rather than arbitrary, and motivated by pedagogy rather than politics. The more power management has with regard to dismissal, the more important it is for them to ensure that justice is not only done, but seen to be done.

Clearly, it would be unethical for managers to discuss with co-workers the specific reasons why a particular colleague was asked to leave. The fairness of the appraisal system would need to be demonstrated in other, more generalised, ways, and its credibility in the eyes of faculty would depend more upon the overall organisational context than upon the specifics of the particular scheme. Simons and Elliott (1989:94) argue, rightly in my view, that if an appraisal system is seen as fair, faculty have no problem with management linking it to disciplinary measures. My point, however, is that whether or not faculty perceive the appraisal system as fair depends not so much upon the actual instruments appraisers use, and the specific procedures they follow, as upon the more general preconceptions appraisees hold about the management of their institution.

According to Poster and Poster (1997), the way in which appraisal is approached will vary according to the school climate and culture. In this way, as was mentioned previously, "... the kind of appraisal system which an instituiton adopts reflects and reveals both the value system and the existing internal structures of the organisation" (Hutchinson, 1997:166). As was noted by the interviewee quoted at length in section 6.5.2, the key issue is not which appraisal procedures are to be used, but whether faculty have confidence that management will apply whichever procedures they adopt fairly and equitably.

7.3 My contribution to the theory and practice of education

7.3.1 The status of my findings

According to Ball (1993:40), "The rigour of any ethnography rests upon the researcher's awareness of what it is possible to say given the nature of the data that was and was not collected". I wholeheartedly agree with Ball's contention and see no reason why it should be limited to ethnographic research. It is therefore encumbent upon me to lay bare the status I would wish to ascribe to the findings and conclusions I have set out above.

Throughout my thesis, I have argued against any positivist belief in a single, objective account of reality, suggesting, instead, that reality consists in the diverse and changing perspectives of many people. I have further argued that my access to these perspectives has been mediated by the socially-constructed nature of data arising from human communication. I have little way of knowing whether what people told me was what they really thought, and absolutely no way of knowing whether what they really thought was true in an ontological sense. "Knowledge is always constructed relative to a framework, to a form of representation, to a cultural code, and to a personal biography ... there is no single, legitimate way to make sense of the world" (Eisner, 1993:54). That is not to say, however, that "*any* story about [a] situation is as good as any other" (Phillips, 1993:60, italics in the original). Stories which exhibit a high-level of criticality and a

thorough openness to scrutiny rightly offer the reader greater 'food for thought' than ill-conceived and badly-written diatribes.

It seems to me that faculty appraisal is a quite emotive topic, and very often, conclusions are drawn on the basis of tenuous evidence. The value of my own research lies in the extent to which its findings are thoroughly grounded in a wealth of reported perceptions, and its methods are thoroughly open to scrutiny. Few other doctoral research projects have undertaken such lengthy and extensive collection of data (two sites, three years, 38 interviews); few other doctoral submissions have included such transparent and detailed description of method ('warts-and-all'). For these reasons, I would argue that my 'story' represents an insightful and illuminating interpretation of a particular social drama, and that my findings are significant, particularly where they run counter to previous studies.

7.3.2 My contribution to the theory and practice of appraisal

The first area of divergence between my own research and much previous literature concerns the purported link between appraisal and better teaching and / or learning. Appraisal in my own study was generally not seen as improving classroom practice nor as facilitating professional development. It would obviously be premature to call for an end to faculty appraisal on the basis that it is failing to achieve its stated aims, but it would certainly be appropriate to examine the rhetoric surrounding appraisal, and conduct further research, exploring why the stated purposes of appraisal (in academic journals and institutional policy documents) are so at odds with the perceived purposes (in the minds of faculty).

Moreover, in much of the literature, the introduction of appraisal into education has been linked to a perceived rise in managerialism, an association that pits public accountability against individual autonomy, and evaluation against professional development. The results of this study, however, suggest that the dichotomy runs deeper, and centres around the extent to which the micro-political machinations present in any institution

are curtailed or propagated by an appraisal system that is subject to greater or lesser legal constraint. The participants in this study reported that they were not opposed to appraisal per se, however evaluative it might be. And neither were they opposed to the removal of unsuitable teachers. What they rejected was the misuse of power in an environment where legal safeguards were minimal.

Accordingly, further research into HE appraisal is needed, particularly in countries other than the UK, USA, Canada, Australia and New Zealand, in order to determine, firstly, how far faculty elsewhere report a similar potential for management misuse of appraisal; and, secondly, how far any such potential is attributed by faculty to particular contextual factors, such as an absence of tenure, a lack of trade unions, and a dearth of government legislation protecting the rights of employees.

7.3.2 My contribution to research methodology

As well as making a modest contribution to the theory and practice of appraisal, my thesis also makes a modest contribution to research methodology. By conducting research consecutively, at two similar sites, I have been able to explore and evaluate alternative research methods. Accordingly, my thesis highlights the possible impact of varying levels of 'intimacy' with regard to insider and outsider research; it details the intricate dilemmas involved in seemingly simple decisions about the timing, location, and recording of interviews; it weighs the advantages and disadvantages of adopting a more or less flexible interview schedule, and a more or less conversational style; it lays bare the assumptions inherent in asking respondents to validate their transcripts; it provides concrete examples of different ways to transcribe audiotapes; and it discusses the ethical issues involved in participant observation.

Naturally, I am not suggesting there can ever be a 'right' way to carry out research based on interview data, but I am suggesting that fellow researchers will be able to make better informed choices about what, exactly, to do in their own situation, having read about the implications and

consequences of the various options I chose at different points in my own research. It is in this way that others will be able to learn from my experience.

PAGE/PAGES EXCLUDED UNDER INSTRUCTION FROM UNIVERSITY

Bibliography

ANDERSON, G. L. and JONES, F. (2000) 'Knowledge generation in educational administration from the inside out: the promise and perils of site-based administration research', *Educational Administration Quarterly,* **36** (3), pp. 428–464.

BACKMAN, C. (1981) 'Social Method and Social Life: appraisal', in BRENNER, M. (ed.) *Social Method and Social Life,* London, Academic Press Limited.

BALL, S. J. (1993) 'Self-doubt and soft data: social and technical trajectories in ethnographic fieldwork', in HAMMERSLEY, M. (ed.) *Educational Research: Volume one: current issues,* London, Paul Chapman Publishing Limited/The Open University.

BASSEY, M. (1996) *Research Intelligence,* Number 58 November 1996.

BASSEY, M. (1999) *Case Study Research in Educational Settings,* Buckingham, Open University Press.

BEER, M. (1986) 'Performance appraisal', in LORSCH, J.W. (ed.) *Handbook of Organizational Behaviour,* New Jersey, Prentice Hall Incorporated.

BELL, L. (1988) 'Staff appraisal and the professional development of teachers', in BELL, L. (ed.) *Appraising Teachers in Schools: a practical guide,* London, Routledge and Kegan Paul.

BENNETT (1999) 'One Drop of Blood: teacher appraisal Mark 2' *Teacher Development*, **3** (3), pp. 411-428.

BILLSBERRY, J. (ed.) (1996) *The Effective Manager,* London, Sage Publications Limited.

BLAXTER, L., HUGHES, C. and TIGHT, M. (1996) *How to Research,* Buckingham, Oxford University Press.

BOSTON, J. (1996) 'Origins and destinations: New Zealand's model of public management and the international transfer of ideas', in WELLER, P. and DAVIS, G. (eds.) *New Ideas, Better Government,* New South Wales, Australia, Allen & Unwin.

BOYATZIS, R. (1998) *Transforming Qualitative Information,* California, Sage Publications Incorporated.

BOYD, E., ALTBACH, P. and WHITELAW, M. (1994) *The Academic Profession: an international perspective,* New Jersey, Carnegie Foundation for the Advancement of Teaching.

BRENNER, M. (1981) 'Patterns of social structure in the research interview', in BRENNER, M. (ed.) *Social Method and Social Life,* London, Academic Press Limited.

BRUNDRETT, M. (1998) 'What lies behind collegiality: legitimation or control?', *Educational Management and Administration,* **26** (3), pp.305-316.

BURGESS, R. (1984) *In the Field: an introduction to field research,* London, Unwin Hyman.

BURGESS, R. (1985) 'The whole truth? Some ethical problems of research in a comprehensive school', in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

BUSH, T. (ed.) (1989) *Managing Education: theory and practice,* Milton Keynes, Open University Press.

CAMPBELL, A. (2002) *Implementing Appraisal: a case study of the implementation of an appraisal system,* Unpublished Ed.D. thesis, The Open University.

CAMPBELL, J. (1989) 'Towards the collegial primary school', in BUSH, T. (ed.) *Managing Education: theory and practice,* Milton Keynes, Open University Press.

CHARMAZ, K (1995) 'Grounded theory' in SMITH, J.A., HARRE, R. and Van LANGENHOVE, L. (eds.) *Rethinking Methods in Psychology,* London, Sage Publications Limited.

CLARKE, J. and NEWMAN J. (1997) *The Managerial State*, London, Sage Publications Incorporated.

COFFEY, A, and ATKINSON, P. (1996) *Making Sense of Qualitative Data: complementary research strategies,* California, Sage Publications Incorporated.

CRESSWELL, J. W. (1996) *Qualitative Inquiry and Research Design: choosing among five traditions,* California, Sage Publications Incorporated.

CULLEN, K. (1997) 'Headteacher appraisal: a view from the inside', *Research Papers in Education,* **12** (2), pp. 177-204.

CURRIE, Jan, and VIDOVICH, L. (2000) 'Privatization and competition policies for Australian universities', *International Journal of Educational Development* **20** (2000), pp. 135-151.

D'APOLLONIA, S. and ABRAMI, P.C. (1997) 'Navigating students' rating of instruction', *American Psychologist,* **52**, pp. 1198-1208.

DEEM, R. (1998) ''New managerialism' and higher education: the management of performance and cultures in universities in the United Kingdom', *International Studies in Sociology of Education,* **8** (2), pp. 47-70.

DEEM. R. (2000) *'New Managerialism' and the Management of UK Universities*, End of Award Report of the Findings of an ESRC Funded Project, Department of Education Research and the Management School, Lancaster University.

DEEM, R. and JOHSON, R. (2000) 'Managerialism and university managers: building new academic communities or disrupting old ones?' in McNAY, I. *Higher Education and its Communities,* Buckingham, SRHE and The Open University Press.

DREVER, E. (1995) *Using Semi-structured Interviews in Small-Scale Research,* Edinburgh, The Scottish Council for Research in Education.

EASTERBY-SMITH, M., THORPE, R. and LOWE, A. (1994) 'The philosophy of research design', in BENNETT, N., GLATTER, R., and LEVACIC, R. (eds.) *Improving Educational Management through research and consultancy,* London, Paul Chapman Publishing Limited.

EDWARDS, A. D. and FURLONG, V.J. (1985) 'Ethnography and theory construction in educational research', in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

EISNER, E. (1993) 'Objectivity in education research', in HAMMERSLEY, M. (ed.) *Educational Research: Volume one: current issues,* London, Paul Chapman Publishing Limited/The Open University.

EISNER, E. (1998) *The Enlightened Eye: qualitative inquiry and the enhancement of educational practice,* New Jersey, Prentice Hall Incorporated.

ELLIOTT, G. (1996) 'Educational management and the crisis of reform in further education', *Journal of Vocational Education and Training,* **48** (1), pp. 5-23.

ELLIOTT, G. and CROSSLEY, M. (1997) 'Contested values in further education', *Educational Management and Administration*, **25** (1), pp.79-92.

ELLIOTT, J. (1983) *Teacher Evaluation and Teaching as a Moral Science,* Cambridge Institution of Education (mimeo).

ENTEMAN, W. (1993) *Managerialism,* Madison, The University of Wiscousin Press.

EVANS, (1995) *An action research enquiry into reflection in action as part of my role as a deputy head,* Unpublished doctoral dissertation, University of Kingston, UK.

EXWORTHY, M. and HALFORD, S. (eds.) (1999) *Professionals and the New Managerialism in the Public Sector,* Buckingham, Open University Press.

FERLIE, E., ASHBURNER, L., FITZGERALD, L. and PETTIGREW, A. (1996) *The New Public Management in Action,* Oxford, Oxford University Press

FIDLER, B. (1989) 'Staff appraisal – theory, concepts and experience in other organisations and problems of adaptation to education', in RICHES, C. and MORGAN, C. (eds.) *Human Resource Management in Education,* Milton Keynes, Open University Press.

FIDLER, B. (1995) 'Staff appraisal and the statutory scheme in England', *School Organisation,* **15** (2), pp. 95-107.

FIDLER, B. and COOPER, R. (eds.) (1992) *Staff Appraisal and Staff Management in Schools and Colleges: a guide to implementation,* Harlow, Longman.

FITZGERALD, T. (2001) 'Potential paradoxes in performance appraisal: emerging issues for New Zealand schools', in MIDDLEWOOD, D. and CARDO, C. (eds.) *Managing Teacher Appraisal and Performance: A comparative approach,* London: Routledge Falmer

FLETCHER, C. (1996) 'Appraisal: an idea whose time has gone?', in BILLSBERRY, J. (ed.) *The Effective Manager,* London, Sage Publications Limited.

FLYNN, R. (1999) 'Managerialism, professionalism and quasi-markets', in EXWORTHY, M. and HALFORD, S. (eds.) *Professionals and the New Managerialism in the Public Sector,* Buckingham, Open University Press.

FOUCAULT, M. (1977) *Discipline and Punish: The birth of the prison,* Harmondsworth, Penguin.

FOUCAULT, M. (1980) 'Two Lectures' in GORDON, C. (ed.) *Power / Knowledge: Selected interviews and other writings 1972 – 1977 by Michel Foucault,* Brighton, Harvester Press.

FOUCAULT, M. (1981) *The History of Sexuality: Vol. 1, the will to knowledge,* Harmondsworth, Penguin.

FOUCAULT, M. (1994) *Ethics: Subjectivity and Truth: Vol. 1, edited by Paul Rabinow,* New York, New York Press.

FRENCH, R. and GREY, C. (1996) *Rethinking Management Education,* London, Sage Publications Limited.

GAGIURI, R. and LITWIN, G. (1968) *Organizational Climate: explorations of a concept,* Boston, Harvard University.

GILHAM, B (2000) *The Research Interview,* London: Continuum.

GITLIN, A., SIEGEL, M. and BORU, K. (1993) 'The politics of method; from Leftist ethnography to educative research', in HAMMERSLEY, M. (ed.) *Educational Research: Volume one: Current issues,* London, Paul Chapman Publishing Limited/The Open University.

GLASER, B. G. and STRAUSS, A. (1967) *The Discovery of Grounded Theory,* Chicago, Aldine.

GLEESON, D. and SHAIN, F. (1999) 'Managing ambiguity: between markets and managerialism – a case study of 'middle' managers in further education', *Sociological Review,* **47** (3), pp 461-490

GRACE, G. (1985) 'Judging teachers: the social and political contexts of teacher evaluation', *British Journal of Sociology of Education,* **6** (1), pp.3-16.

GREENWALD, A.G. (1997) 'Validity concerns and usefulness of student ratings of instruction', *American Psychologist,* **52**, pp. 1182-1186.

GREENWALD, A.G. and GILLMORE, G.M. (1997) 'Grading leniency is a removable contaminant of student ratings', *American Psychologist,* **52**, pp. 1209-1217.

GRIFFIN, C. (1985) 'Qualitative methods and cultural analysis' in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

GRIFFITHS, G. (1985) 'Doubts, Dilemmas and Diary-Keeping: Some Reflections on Teacher-Based Research' in BURGESS, R. (ed.) *Issues in Educational Research: Qualitative Methods*, London, The Falmer Press, pp. 197-215.

GUNTER, H. (1996) 'Appraisal and the school as a learning organisation', *School Organisation*, **16** (1), pp. 89-100.

GUNTER, H. (1999) 'Adding values: the role of the LEA appraisal co-ordinators in the implementation of the teacher appraisal 1991-1996', *Educational Management and Administration,* **27** (4), pp.375-387.

HALL, V. (1996) 'Managing Staff' in FIDLER, B., RUSSELL, S. and SIMPKINS, T. (eds.) *Choices of Self-managing Schools: autonomy and accountability,* London, Paul Chapman.

HAMMERSLEY, M. (ed.) (1993) *Educational Research: Volume One; Current issues,* London, Paul Chapman Publishing Limited/The Open University.

HANCOCK, N. and HELLAWELL, D. (1998) 'Primary school headteachers' attempts to maintain a person-centred managerial style: the evidence from their appraisals", *Teacher Development,* **2** (4), pp. 251-264.

HARGREAVES, D. (1996) 'Educational research and evidence-based educational practice – a response to critics' in BASSEY, M. *Research Intelligence,* Number 58 November 1996 pp.12-16.

HARRISSON, N. (1994) 'Managing an aspect of pastoral care in a primary school', in CRAWFORD, M., KYDD, L. and PARKER, S. (eds.) *Educational Management in Action,* London, Paul Chapman Publishing Limited.

HARTLEY, D. (1997) 'The new managerialism in education: a mission impossible?', *Cambridge Journal of Education,* **27** (1), pp. 47-57.

HAWKINS, B.S.R. (1990) *The Management of Staff Development in a Contracting Education Service,* Unpublished Ph.D. thesis, Birmingham Polytechnic.

HELLAWELL, D. (1991) *Primary Headteacher Appraisal: A research study of the attitudes of headteachers,* Sheffield Papers in Education Management 89, Sheffield City Polytechnic, Centre for Education Management and Administration, Sheffield.

HELLAWELL, D. and HANCOCK, N. (2001) 'A case study of the changing role of the academic middle manager in higher education: between hierarchical control and collegiality', *Research Papers in Education,* **16** (2), pp. 1-15.

HOCKEY, J. (1993) 'Research methods – researching peers and familiar settings', *Research Papers in Education,* **8** (2), pp. 199-225.

HODKINSON, P. (1998) 'Naiveté and bias in educational research: the Tooley report', *Research Intelligence,* Number 65, pp. 16-17.

HOFSTEDE, G. (1991) *Cultures and Organisations: software of the mind,* Maidenhead, McGraw-Hill Book Company (UK) Limited.

HOFSTEDE, G. (1996) 'The cultural relativity of organisational practices and theories', in BILLSBERRY, J. (ed.) *The Effective Manager,* London, Sage Publications Limited.

HOLLIDAY, A. (2002) 'Collecting and writing about data in qualitative research', conference given at Dubai Women's Higher College of Technology, 17 March, 2002.

HOLLOWAY, D. (1999) 'The audit commission, managerialism and the further education sector', *Journal of Vocational Education and Training,* **51** (2), pp. 229 – 243.

HOUSE, E. and LAPAN, S. (1989) 'Teacher Appraisal' in SIMONS, H. and ELLIOTT, J. (eds.) *Rethinking Appraisal and Assessment,* Milton Keynes, Open University Press.

HUTCHINSON, B. (1997) 'Appraising appraisal: some tensions and some possibilities', in KYDD, L., CRAWFORD, M. and RICHES, C. (eds.) *Professional Development for Educational Management,* Buckingham, Open University Press.

INGLIS, F. (1989) 'Managerialism and morality', in CARR, W. (ed.) *Quality in Teaching,* London, The Falmer Press.

KO, E. (2001) Performance Appraisal in Higher Education: the impact of new managerialism on the motivation of academic staff, University of Bristol, unpublished Ed.D. thesis.

KYDD, L. (1997) 'Teaching professionalism and managerialism', in KYDD, L., CRAWFORD, M. and RICHES, C. (eds.) *Professional Development for Educational Management,* Buckingham, Open University Press.

KYRIACOU, C. (1992) 'Observation of classroom teaching', in FIDLER, B. and COOPER, R. (eds.) *Staff Appraisal and Staff Management in Schools and Colleges,* Harlow, Longman.

KYRIACOU, C. (1995) 'An evaluation of teacher appraisal in schools within one Local Education Authority', *School Organisation,* **15** (2), pp. 109-116.

LINCOLN, Y.S. and GUBA, E.G. (1985) *Naturalistic Inquiry,* Beverley Hills, CA, Sage.

LINGARD, B., LADWIG, J. and LUKE, A. (1998) 'School effects in postmodern conditions', in SLEE, R., WEINER, G and TOMLINSON, S. (eds.) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements,* Lewes, Falmer Press

LOGAN, T. (1984) 'Learning through interviewing', in SCHOSLAK, J. and LOGAN, T. (eds.) *Pupil Perspectives,* London, Croom Helm.

LOVRICH, N.P. (1990) 'Performance Appraisal' in HAYS, S.W. and KEARNEY, R.C. (eds.) *Public Personnel Administration: problems and prospects,* New Jersey, Prentice Hall.

LUMBY, J. and TOMLINSON, H. (2000) 'Principals speaking: managerialism and leadership in further education', *Research in Post-compulsory Education,* **5** (2), pp. 139–151.

MAGENNIS, S. (1993) 'Appraisal systems and their contribution to quality in teaching', in ELLIS, R. (ed.) *Quality Assurance in University Teaching,* Buckingham, The Open University Press.

MAGGS-RAPPORT, F. (2000) 'Combining methodological approaches in research: ethnography and interpretive phenomenology', *Journal of Advanced Nursing*, **31** (1), pp. 219-225.

MARSH, H.W. and ROCHE, L.A. (1997) 'Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility', *American Psychologist,* **52**, pp. 1187-1197.

MARUYAMA, G. and DENO, S. (1992) *Research in Educational Settings, Sage Applied Social Research Methods Series,* **29**, California, Sage Publications Incorporated.

MEASOR, L. and WOODS, P. (1991) 'Breakthroughs and blockages in ethnographic research: contrasting experiences during the C*hanging Schools* project', in WALFORD, G. (ed.) *Doing Education Research,* London, Routledge.

MEYER, H-D. (2002) 'The new managerialism in education management: Corporatization or organizational learning?', *Journal of Educational Administration,* **40** (6), pp. 534-552.

MILES, M. B. and HUBERMAN, A. M. (1984) *Qualitative Data Analysis: a handbook of new methods,* California, Sage Publications Incorporated.

MORRIS, B. (1991) 'Schoolteacher appraisal: reflections on recent history', *Educational Management and Administration*, **19** (3), pp. 166-171.

MORRISON, K. and RIDLEY, K. (1989) 'Ideological context in curriculum planning', in PREEDY, M. (ed.) *Approaches to Curriculum Management,* Buckingham, Open University Press.

MORTIMORE, P. and MORTIMORE, J. (1991) 'Teacher appraisal: back to the future', *School Organisation,* **11** (2), pp. 125-143.

MURRAY, R. and HOLMES, S. (1997) 'Partnership in Staff Development: an institutional case study', *Studies in Higher Education,* **22** (1).

NIAS, J. (1991) 'Primary teachers talking: a reflexive account of longitudinal research', in WALFORD, G. (ed.) *Doing Education Research,* London, Routledge.

NUTTALL, D. (1986) 'What can we learn from research on teaching and appraisal?', *Report of a Conference organised by the British Educational Research Association,* Birmingham: British Research Association.

OAKLEY, A. (1981) 'Interviewing women', in ROBERTS, H. (ed.) *Doing Feminist Research,* London, Routledge and Kegan Paul.

OGBANNA, E. (1996) 'Managing organisational culture: fantasy or reality?', in BILLSBERRY, J. (ed.) *The Effective Manager,* London, Sage Publications Limited.

OPEN UNIVERSITY (1989) *E818: Management in Education,* Milton Keynes, Open University Press.

PATTON, M. (1999) 'Enhancing the quality and credibility of qualitative analysis', *Health Services Research,* **34** (5), pp. 1189-1208.

PEAKER, G. (1986) 'Teacher management and appraisal in two school systems in the southern USA', *Journal of Education for Teaching,* **12** (1), pp. 77-83.

PETERS, M., MARSHALL, J. and FITZSIMONS, P. (2000) 'Managerialism and educational policy in a global context: Foucault, Neoliberalism and the doctrine of self-management', in BURBULES, N. and TORRES, C. (eds.) *Globalization and Education: cultural perspective,* New York, Routledge.

PHILLIPS, D.C. (1993) 'Subjectivity and objectivity: an objective inquiry' in in HAMMERSLEY, M. (ed.) *Educational Research: Volume one: current issues,* London, Paul Chapman Publishing Limited/The Open University.

PLATT, J. (1981) "On interviewing one's peers", *The British Journal of Sociology,* **32** (1), pp.75-91.

POLLARD, A. (1985) 'Opportunities and difficulties of a teacher-ethnographer: a personal account', in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

POLLITT, C. (1987) 'The politics of performance assessment: lessons for higher education', *Studies in Higher Education,* **12** (1), pp. 87-98.

POLLITT, C. (1990) *Managerialism and the Public Services: The Anglo-American experience,* Oxford, Blackwell.

POLLITT, C. (1993) *Managerialism and the Public Services (Second Edition),* Oxford, Blackwell.

POSTER, C. and POSTER, D. (1997) 'The nature of appraisal', in KYDD, L., CRAWFORD, M. and RICHES, C. (eds.) *Professional Development for Educational Management,* Buckingham, Open University Press.

POWNEY, J. (1991a) 'The appraisal of teachers in middle management roles', *Research Papers in Education,* **6** (3), pp.171-195.

POWNEY, J. (1991b) 'Teacher appraisal: the case for a developmental approach', *Educational Research,* **33** (2) pp. 83-91.

POWNEY, J. and WATTS, M. (1987) *Interviewing in Educational Research,* London, Routledge and Kegan Paul Limited.

RANDLE, K. and BRADY, N. (1997a) 'Further Education and the new managerialism', *The Journal of Further and Higher Education,* **21** (2), pp. 229-239.

RANDLE, K. and BRADY, N. (1997b) 'Managerialism and professionalism in the 'Cinderella Service'', *Journal of Vocational Education and Training,* **29** (1), pp. 121–139.

REED, M. (2002) 'New managerialism and new forms of organisational control in UK higher education', *Management Research News,* **25** (8-10), pp.158-163.

RICHARDS, J. (1999) *Introducing a New Appraisal Scheme: a case study of a tertiary college in the Middle East,* (unpublished M.Ed. dissertation, University of Sheffield.)

ROBSON, C. (1993) *Real World Research: a resource for social scientists and practitioner-researchers,* Oxford, Blackwell Publishers Limited.

RUTHERFORD, D. (1988) 'Performance appraisal: a survey of academic staff opinion', *Studies in Higher Education,* **13** (1), pp. 89-100.

SALE, J. (1997) 'Appraisal Matters', *Managing Schools Today,* **6** (7), pp. 28-29.

SCOFIELD, J. (1993), 'Increasing the generalizability of qualitative research', in HAMMERSLEY, M. (ed.) *Educational Research: Volume one: current issues,* London: Paul Chapman Publishing Limited/The Open University.

SCOTT, S. (1985) 'Working through the contradictions in researching postgraduate education', in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

SHELLEY, S. (1999) 'Conflicts and contrasts: managerialism and the labour processes in universities', *Working Paper Series,* Business School, University of Hertford.

SILVERMAN, D. (1993) *Interpreting Qualitative Data: methods, for analysing talk, text and interaction,* London, Sage Publications.

SILVERMAN, D. (2000) *Doing Qualitative Research: a practical handbook,* London, Sage Publications.

SIMKINS, T. (2000) 'Education reform and managerialism: comparing the experience of schools and colleges', *Journal of Education Policy,* **15** (2), pp. 317- 332.

SIMONS, H. and ELLIOTT, J. (eds.) (1989) *Rethinking Appraisal and Assessment,* Milton Keynes, Open University Press.

SMITH, J.A. (1995) 'Semi-structured interviews and qualitative analysis' in SMITH, J.A., HARRE, R. and Van LANGENHOVE, L. (eds.) *Rethinking Methods in Psychology,* London, Sage Publications Limited.

SOFAER, S. (1999) 'Qualitative methods: What are they and why use them?', *Health Services Research,* **34** (5), pp. 1101-1118.

SPARKES, A.C. (1995) 'Writing people: reflections on the dual crisis of representation and legitimation in qualitative inquiry' in *Quest,* 47.

STAKE, R. E. (1989) 'The evaluation of teaching', in SIMONS, H. and ELLIOTT, J. (eds.) *Rethinking Appraisal and Assessment,* Milton Keynes, Open University Press.

STAKE, R. E. (1995) *The Art of Case Study Research,* London, Sage.
STENHOUSE, L. (1985) 'A note on case study and educational practice', in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

STRAUSS, A. and CORBIN, J. (1990) *Basics of Qualitative Research: grounded theory procedure and techniques,* California, Sage Publications Limited.

STROMQUIST, N. (2000) 'On truth, voice and qualitative research', *International Journal in Qualitative Studies in Education,* **13** (20), pp.139-152.

TAYLOR, H. (1997) 'A case study in the development of a staff appraisal system', *Journal of Further and Higher Education,* **21** (1), pp. 55-61.

TOMLINSON, B. (1998) 'A response to Brundrett?', *Educational Management and Administration,* **26** (3), pp. 317-319.

TOOLEY, J. and DARBY, D. (1998) *Education Research - a critique: a survey of published educational research*, London: Office For Standards In Education.

TROMAN, G. (1996) 'Models of the 'good' teacher: defining and redefining teacher quality' in WOODS, P. (ed.) *Contemporary Issues in Teaching and Learning,* Buckingham, The Open University.

TROW, M. (1994) 'Managerialism and the academic profession: the case of England', *Higher Education Policy*, **17** (2), pp. 11-18.

TROW, M. (1996) 'On the accountability of Higher Education in the United States' in BOWEN, W. and SHAPIRO, H. (eds.) *Universities and Their Leadership,* Princeton, Princetown University Press.

TROWLER, P. (1998a) 'What managerialists forget: higher education credit frameworks and managerialist ideology', *International Studies in Sociology of Education,* **8** (1), pp. 91-109.

TROWLER, P. (1998b) *Academics Responding to Change: new higher education frameworks and academic cultures,* Buckingham, SRHE and Open University Press

TURNER, G, and CLIFT, P. (1988) *Studies in Teacher Appraisal,* Lewes, The Falmer Press

WALKER, R. (1980) 'The Conduct of Educational Case Studies: Ethics, Theory and Procedures', in DOCKRELL, W.B. and HAMILTON (eds.) *Rethinking Educational Research,* London: Hodder and Stoughton.

WALLACE, M. (1988) 'Towards a collegiate approach to curriculum management in primary and middle schools', *School Organisation,* **8** (1), pp. 25-34.

WALSH, K. (1988) 'Appraising the teachers: professionalism and control', in DALE R, et al. (eds.) *Frameworks for Teaching,* Milton Keynes, The Open University Press.

WEBB, R. and VULLIAMY, G. (1996) 'A deluge of directives: conflict between collegiality and managerialism in the post ERA primary school', *British Educational Research Journal*, **22** (4), pp. 441-458.

WEICK, K. (2001) *Making Sense of Organisations,* Oxford, Blackwell

WILSON, R. (1998) 'New research casts doubt on value of student evaluations of professors' *The Chronicle of Higher Education,* January 16, 1998, accessed on-line at http://chronicle.com/data/articles.dir/art-44.dir/issue-19.dir/19a01201.htm

WINTER, R. (1989a) 'Teacher appraisal and the development of professional knowledge', in CARR, W. (ed.) Q*uality in Teaching,* London, Falmer Press.

WINTER, R. (1989b) 'Problems in teacher appraisal: an action-research solution?', in SIMONS, H. and ELLIOTT, J. (eds.) *Rethinking Appraisal and Assessment,* Milton Keynes, Open University Press.

WINTER, R., TAYLOR, T. and SARROS, J. (2000), 'Trouble at mill: Quality of academic worklife issues with a comprehensive Australian university, *Studies in Higher Education,* **25** (3) pp 279-295.

WOODS, P. (1985) 'Ethnography and theory construction in educational research', in BURGESS, R. (ed.) *Field Methods in the Study of Education,* Lewes, Falmer Press.

WOODS, P and JEFFREY, R. (1996) 'A new professional discourse? Adjusting to managerialism', in WOODS, P. (ed.) *Contemporary Issues in Teaching and Learning,* Buckingham, The Open University.

WRAGG, E.C. (1984) *Classroom Teaching Skills,* London, Croom Helm.

WRAGG, E., WIKELEY, F., WRAGG, C. and HAYNES, G. (1996) *Teacher Appraisal Observed,* London, Routledge.

YIN, R.K. (1993) *Applications of Case Study Research,* London, Sage.